

SUSANNE LISA SCHMIDT

CHUNKS FÜR DAF

Eine Lernerkorpusanalyse zur Identifikation und Klassifizierung
von Chunks für DaF

Curitiba
2016

SUSANNE LISA SCHMIDT

CHUNKS FÜR DAF

Eine Lernerkorpusanalyse zur Identifikation und Klassifizierung
von Chunks für DaF

Dissertação apresentada como requisito
parcial à obtenção do título de mestre
junto à área de concentração em Estudos
Linguísticos do Programa de Pós-
Graduação em Letras da Universidade
Federal do Paraná.

Linha de pesquisa: Alemão como Língua
Estrangeira

Orientador: Prof. Dr. Erwin Tschirner
Coorientadora: Prof. Dr. Ruth Bohunovsky

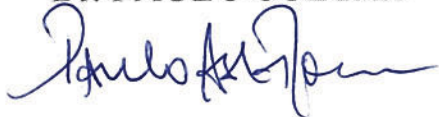
Curitiba
2016

Ata septingentésima octogésima primeira, referente à sessão pública de defesa de dissertação para a obtenção de título de mestre a que se submeteu a mestranda **SUSANNE LISA SCHMIDT**. No dia vinte e dois de fevereiro de dois mil e dezessete, às treze horas, na Sala 1012, 10.º andar, no Setor de Ciências Humanas da Universidade Federal do Paraná, foram instalados os trabalhos da Banca Examinadora, constituída pelos seguintes Professores Doutores: ERWIN TSCHIRNER, PAULO SOETHE e RUTH BOHUNOVSKY, designados pelo Colegiado do Curso de Pós-Graduação em Letras, para a sessão pública de defesa de dissertação intitulada “CHUNKS FUER DAF: EINE LERNERKORPUSANALYSE ZUR IDENTIFIKATION UND KLASSIFIZIERUNG VON CHUNKS FUER DAF”, apresentada por **SUSANNE LISA SCHMIDT**. A sessão teve início com a apresentação oral da mestranda sobre o estudo desenvolvido. Logo após, a senhora presidente dos trabalhos concedeu a palavra a cada um dos examinadores para as suas arguições. Em seguida, a candidata apresentou sua defesa. Na sequência, a Professora RUTH BOHUNOVSKY retomou a palavra para as considerações finais. Na continuação, a Banca Examinadora, reunida sigilosamente, decidiu pela aprovação da candidata. Em seguida, a senhora Presidente declarou **APROVADA** a candidata, que recebeu o título de **Mestre em Letras**, área de concentração **Estudos Linguísticos**. A versão final da dissertação deverá ser encaminhada à Coordenação em até 60 dias. Encerrada a sessão, lavrou-se a presente ata, que vai assinada pela Banca Examinadora e pela candidata. Feita em Curitiba, no dia vinte e dois de fevereiro de dois mil e dezessete.

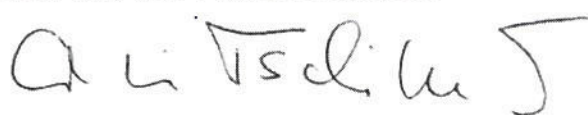
Dr.^a RUTH BOHUNOVSKY



Dr. PAULO SOETHE



Dr. ERWIN TSCHIRNER



SUSANNE LISA SCHMIDT

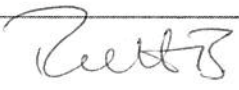
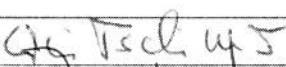
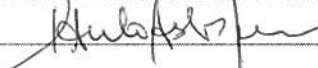


PARECER

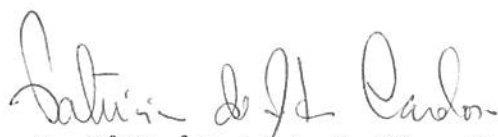
Defesa de dissertação de mestrado de SUSANNE LISA SCHMIDT para obtenção do título de **Mestre em Letras**.

Os abaixo - assinados ERWIN TSCHIRNER, PAULO SOETHE e RUTH BOHUNOVSKY arguíram, nesta data, a candidata, que apresentou a dissertação "**CHUNKS FUER DAF: EINE LERNERKORPUSANALYSE ZUR IDENTIFIKATION UND KLASSIFIZIERUNG VON CHUNKS FUER DAF**".

Procedida a arguição segundo o protocolo que foi aprovado pelo Colegiado do Curso, a Banca é de parecer que a candidata está apta ao título de **Mestre em Letras**, conforme especificações abaixo:

Banca	Assinatura	APROVADA Não APROVADA
Dr. ^a RUTH BOHUNOVSKY (Presidente)		aprovada
Dr. ERWIN TSCHIRNER		Aprovada
Dr. PAULO SOETHE		APROVADA

Curitiba, 22 de fevereiro de 2017.



Prof.^a Dr.^a Patrícia da Silva Cardoso
Coordenadora

Catálogo na publicação
Mariluci Zanela – CRB 9/1233
Biblioteca de Ciências Humanas e Educação - UFPR

Schmidt, Susanne Lisa

Chunks für daf: eine lernerkorporusanalyse zur identifikation und klassifizierung von Chunks für Daf / Susanne Lisa Schmidt – Curitiba, 2016.

174 f.; 29 cm.

Orientador: Erwin Tschirner

Coorientadora: Ruth Bohunovsky

Dissertação (Mestrado em Letras) – Setor de Ciências Humanas da Universidade Federal do Paraná.

1. Psicolinguística. 2. Linguística de corpus. 3. Língua alemã – Estudo de ensino. I. Título.

CDD 435

Este trabalho toma como ponto de partida a premissa de que falantes nativos e não-nativos adquirem a língua, ou seja, um conhecimento implícito das estruturas gramaticais enquanto gravam *chunks* – sequências de palavras que formam uma unidade semântica e são armazenadas de forma holística. Esta perspectiva rompe com teorias precedentes na medida em que expressões não são mais classificadas entre regulares e irregulares mas podem ser organizadas ao longo de um continuum léxico-sintático de mais ou menos abstratas, abandonando a separação do léxico e da sintaxe. À luz dessas reflexões e de suas implicações para o ensino de línguas estrangeiras, a análise tem por objetivo a extração e avaliação de *chunks* frequentes num corpus contendo produções nativas e não-nativas em alemão. Além disso, o trabalho investiga em que aspectos o uso dos *chunks* nativos e não-nativos diverge. Uma avaliação dos resultados é apresentada.

Palavras-chave chunks; psicolingüística; lingüística de corpus, frequência; ensino de línguas estrangeiras

Usage-based theories of language hold that native and non-native speakers acquire the language, that is an inventory of *constructions* – conventionalized form-meaning mappings – by memorizing *chunks* – formulaic sequences that form a semantic unit and are stored holistically. This approach breaks with previous theories of language insofar as expressions are no longer classified as regular or irregular but can be arranged on a syntax-lexicon continuum ranging from the lexically-specific to the fully abstract; thus dismissing the separation of what has been traditionally designated as lexicon and grammar. Considering these findings and their implications for language instruction, the aim of this study is to identify, analyze and compare frequent native and non-native chunks in a learner corpus of German. Findings of the study as well as implications for future work will be discussed.

Keywords chunks; psycholinguistics; corpus data; frequency, language instruction

INHALTSVERZEICHNIS

TABELLEN- UND ABBILDUNGSVERZEICHNIS	II
1. EINLEITUNG	5
2. CHUNKS – BEGRIFFSERKLÄRUNG	8
2.1 PSYCHOLINGUISTIK	8
2.2 PHRASEOLOGIE	10
2.3 SPRACHTHEORETISCHE VERORTUNG	16
2.3.1 KONSTRUKTIONSGRAMMATIK	18
2.3.2 KONSTRUKTIKON	21
3. CHUNKS IM ERST- UND ZWEITSPRACHERWERB	23
3.1 CHUNKS IM ERSTSPRACHERWERB	23
3.1.1 NATIVISTISCH VS. GEBRAUCHSBASIERT	24
3.1.2 FAKTOREN IM GEBRAUCHSBASIERTEN ERSTSPRACHERWERB	26
3.1.2.1 Frequenz	26
3.1.2.2 Kategorisierung	28
3.1.2.3 Zuverlässigkeit und Form	30
3.1.2.4 Cross-linguistische Variation	31
3.2 CHUNKS IM ZWEITSPRACHERWERB	33
3.2.1 FAKTOREN IM GEBRAUCHSBASIERTEN ZWEITSPRACHERWERB	35
3.2.1.1 Limitierter Input	35
3.2.1.2 Transfer	36
3.2.2 IMPLIKATIONEN FÜR DEN GESTEUERTEN ZWEITSPRACHERWERB	38
4. ANSÄTZE ZUR VERMITTLUNG VON CHUNKS	39
4.1 THE LEXICAL APPROACH	39
4.2 DER CHUNK-UND-KONSTRUKTIONEN-ANSATZ	42
5. KORPUSLINGUISTIK	43
5.1 LERNERKORPORA	45
5.2 KORPUSANALYTISCHE METHODEN ZUR IDENTIFIKATION VON CHUNKS	46
5.2.1 FREQUENZ UND DISPERSION	46
5.2.2 ASSOZIATIONSMAßE	47
6. ANALYSE	48
6.1 ZIELSETZUNGEN DER ANALYSE	48
6.2 DATENGRUNDLAGE	49
6.2.1 DAS LERNERKORPUS FALKO	49
6.2.2 METADATEN DER TEILNEHMENDEN	50
6.3 DATENERHEBUNG	51
6.4 AUSWERTUNG	53
6.4.1 EIGENSCHAFTEN VON L2-CHUNKS	54
6.4.1.1 Dreigliedrige Chunks	55
6.4.1.2 Viergliedrige Chunks	57
6.4.1.3 Assoziative Chunks	58
6.4.1.4 Erweiterte L2-Chunks	60
6.4.2 VERGLEICH L1 – L2	62
6.5 DISKUSSION	64
7. AUSBLICK	66
LITERATURVERZEICHNIS	III
ANHANG	X

TABELLEN- UND ABBILDUNGSVERZEICHNIS

Tabellenverzeichnis

Tabelle 1. Typologie formelhafter Wendungen	15
Tabelle 2. Subkorpora des Lernerkorpus Falko	49
Tabelle 3. Verteilung der am besten beherrschten Muttersprachen	51
Tabelle 4. Auswahl dreigliedriger Chunks	54
Tabelle 5. Auswahl viergliedriger Chunks	55
Tabelle 6. Viergliedrige assoziative Chunks	59
Tabelle 7. Chunk / [Dd]ie/ . / [Mm]ehrheit/ . "der" . pos="NN" (16)	60
Tabelle 8. Die frequentesten dreigliedrigen L1-Chunks im Vergleich zu L2	62
Tabelle 9. Die assoziativsten dreigliedrigen L1-Chunks im Vergleich zu L2	62

Abbildungsverzeichnis

Abbildung 1. Konstruktionen als Form-Bedeutungspaare	19
Abbildung 2. (Ausschnitt aus einem) Netzwerk formseitig verbundener Konstruktionen	21
Abbildung 3. Komponenten des Chunk-und-Konstruktionen-Ansatzes	43
Abbildung 4. Ausschnitt der häufigsten Trigramme im Lernerkorpus	52
Abbildung 5. Ausschnitt assoziativer Trigramme im Vergleichskorpus	53
Abbildung 6. Bedeutungsdimensionen des Schemas <i>in d X</i>	56

1. EINLEITUNG

Grundlegend für das Forschungsinteresse der vorliegenden Arbeit ist die Erkenntnis, dass Vorgeprägtheit und Musterhaftigkeit einen weitaus höheren Stellenwert im Sprachsystem sowie -gebrauch einnehmen als traditionell angenommen. Dabei wird *formelhafte Sprache* zum Untersuchungsgegenstand unterschiedlicher Disziplinen und Bereiche. So beschäftigen sich u.a. Phraseologie (Burger 2015; Steyer 2013), kognitive Linguistik (Robinson/Ellis 2008), Theorien des Erst- (Tomasello 2003; 2006) und Zweitspracherwerbs (Ellis 2007) sowie Grammatiktheorien (Goldberg 1995) mit dem Vorgefertigten in der Sprache. Die Reichweite des Phänomens wird zudem anhand einer neuen Buchreihe *Formelhafte Sprache / Formulaic Language* (Filatkina et al.: n.d.) deutlich, für die aktuell Studien gesammelt werden, die das Formelhafte in Bereichen, die von „Lexikon und Grammatik“ bis zu „Kognition und kulturelles Gedächtnis“ reichen, untersuchen.

Der Abruf ganzheitlich gespeicherter, vorgefertigter Sequenzen bietet dabei eine Antwort auf die „zwei sprachtheoretischen Rätsel“ *nativelike selection* und *nativelike fluency* (Pawley/Syder 1983), d.h. die Fähigkeit von erwachsenen Muttersprachlern, sich natürlich und in spontanen Diskursen erstaunlich schnell und ohne Sprechpausen auszudrücken. Pawley und Syder (vgl. ebd.: 213) gehen davon aus, dass die Anzahl an abrufbaren Konstrukten mehrere hunderttausend umfasst. Ertan und Warren (2000) untersuchten 19 Auszüge aus Korpora gesprochener und geschriebener Sprache für das Englische und fanden heraus, dass vorgefertigte Sequenzen zwischen 40 und 60 Prozent der Diskurse bzw. Texte ausmachten¹.

Die Erkenntnis, dass Formelhaftigkeit im Sprachsystem sowie in der Sprachverwendung dominiert, ist eng mit dem Aufschwung der Korpuslinguistik (Sinclair 1987) verknüpft, die es erstmals ermöglichte, große Textsammlungen schriftlicher oder gesprochener Sprache maschinell zu durchsuchen und so Muster im Sprachgebrauch zu identifizieren. Im Zuge dessen erweiterte auch die Phraseologie, die sich klassischerweise mit auffälligen, idiomatischen Ausdrücken beschäftigte, ihren Untersuchungsgegenstand. Dies ist für die Praxis des Deutsch als Fremdsprache-Unterrichts insofern relevant, als dass

[...] die in Lehrwerken und didaktischen Materialien für Deutsch als Fremdsprache aufgeführten Wortverbindungen eher untypisch oder gar archaisch anmuten. Dieses oft bestehende Missverständnis zwischen dem gesammelten Material bzw. den bevorzugt behandel-

¹ Es muss jedoch beachtet werden, dass unter anderem die Definition, also was gilt als formelhafte Sequenz, sowie Korpusdesign und Stichprobe einen enormen Einfluss darauf haben, welche Anteile und welche Arten formelhafter Sprache gefunden werden (vgl. O'Donnell et al. 2013: 101).

ten Beschreibungsobjekten einerseits und den tatsächlich aktuell üblichen Wortverbindungen andererseits hat eine objektive Ursache, die bereits angeklungen ist: Man hat sich in vortechnologischer Zeit naturgemäß eher mit den Objekten beschäftigt, die auffällig waren, z.B. auf Grund ihrer opaken Bedeutung, ihrer Idiosynkrasie, ihrer Ungewöhnlichkeit in Form und Bedeutung und auch auf Grund ihres seltenen Vorkommens. Dies spiegelte und spiegelt sich dann auch in der lexikographischen und didaktischen Praxis wider.

(Steyer 2013: 76)

Für den Fremdsprachenlernenden² stellen formelhafte Sequenzen dabei eine besondere Herausforderung dar, vor allem hochassoziative Mehrworteinheiten, deren Einzelteile nicht oft alleine vorkommen, werden zu selten von Lernenden einer zweiten oder weiteren Sprache (L2) genutzt (Simpson-Vlach/Ellis 2009; Ellis et al. 2008; Durrant/Schmitt 2009). Rezektiv haben L2-Lernende Schwierigkeiten, Chunks als solche zu erkennen, da sie primär einzelne Wörter fokussieren (vgl. Wray 2002: 200). Gleichzeitig stellt das Lernen von Chunks ein großes Potenzial für den Fremdsprachenunterricht dar, da es Lernende wie Muttersprachler befähigt, schneller und „natürlicher“ zu sprechen und sie grammatische Strukturen verwenden lässt, die ihre Lernergrammatik überschreiten. Ferner konnte nachgewiesen werden, dass Lernende wie auch Muttersprachler abstrakte Regularitäten aus dem Sprachangebot ableiten (Gries/Wulff 2005), d.h. im gesteuerten Zweitspracherwerb kann ein strukturierter Input mit Chunks variierender Komplexität den Aufbau implizit grammatischen Wissens bei den Lernenden fördern (Handwerker 2008).

Der Ansatz stellt insofern einen Bruch mit früheren Sprach(erwerbs)theorien dar, als dass Äußerungen nicht mehr in regelmäßig und unregelmäßig eingeteilt, sondern diese einem Lexikon-Syntax-Kontinuum von mehr oder minder abstrakt zugeordnet werden. Zudem wird Sprache erschöpfend in Form von Form-Bedeutungspaaren beschrieben, sodass Syntax, Semantik und Pragmatik nicht mehr unabhängig voneinander untersucht werden.

Davon ausgehend, dass mehr oder minder abstrakte Chunks als Basiseinheiten im Spracherwerb und der Sprachverwendung fungieren und so ein geeigneteres Format zur Vermittlung im Fremdsprachenunterricht als einzelne Wörter darstellen, soll in dieser Arbeit folgenden Fragen nachgegangen werden:

- (i) Welche Chunks werden von Lernenden auf fortgeschrittenen Niveaus verwendet? bzw. Welche Eigenschaften weisen diese Chunks auf?
- (ii) Inwieweit unterscheidet sich der Gebrauch formelhafter Chunks von Lernenden und Muttersprachlern?

² I.d.R. wurden neutrale Formen verwendet, um die sprachliche Gleichstellung aller Geschlechter sichtbar zu machen. Wenn dies aus Gründen der besseren Lesbarkeit nicht erfolgte, gelten die Personenbezeichnungen gleichwohl für alle Geschlechter.

Zum einen sollen also häufige, rekurrente L2-Chunks empirisch erfasst und im Anschluss unter Berücksichtigung struktureller, semantischer und pragmatischer Aspekte beschrieben werden. Zum anderen sollen extrahierte L1-Chunks vergleichend für L1 (Erstsprache) und L2 untersucht werden, um Gebrauchspräferenzen offenzulegen.

Die Arbeit geht dabei in folgenden Schritten vor: Zunächst soll auf den Untersuchungsgegenstand *Chunk* genauer eingegangen werden, wobei Definitionen und Erkenntnisse aus der Psycholinguistik, der Phraseologie und der Konstruktionsgrammatik zusammengetragen und zueinander in Bezug gesetzt werden. Kapitel drei beleuchtet im Anschluss die Rolle von *Chunking* im Erst- und Zweitspracherwerb sowie kognitive Prozesse und Inputfaktoren, die das Erlernen von Chunks beeinflussen. Kapitel vier widmet sich Ansätzen, die das Chunk-Lernen für die praktische Anwendung im Fremdsprachenunterricht nutzbar machen wollen. In Kapitel fünf werden korpusanalytische Methoden, die sich zur Extraktion von Chunks eignen, vorgestellt. In Kapitel sechs wird das Falko-Korpus bzw. dessen Subkorpora, die als Datengrundlage dienen, beschrieben. Anschließend werden die häufigsten Drei- und Vierworteinheiten im Lerner- sowie im muttersprachlichen Vergleichskorpus ermittelt, strukturiert und miteinander verglichen. Zudem werden Konsequenzen, die sich daraus hinsichtlich Wortschatz- und Grammatikvermittlung ergeben, im Ansatz aufgezeigt. Kapitel sieben fasst abschließend die Resultate der Arbeit zusammen, beleuchtet Grenzen der durchgeführten Analyse und formuliert Desiderate für zukünftige Arbeiten.

2. CHUNKS – BEGRIFFSERKLÄRUNG

Nachfolgend soll sich dem Phänomen *Chunking* aus psycholinguistischer, phraseologischer und sprachtheoretischer Sicht genähert werden, wobei die Psycholinguistik von der holistischen Verarbeitung und Speicherung von Informationsbündeln ausgeht und die Phraseologie weitere Eigenschaften von Chunks, wie Polylexikalität, Festigkeit und Idiomatizität beschreibt. Darüber hinaus nehmen Chunks bzw. Konstruktionen eine zentrale Rolle in neueren Sprachtheorien ein, wo sie als Basiseinheiten von Sprache fungieren und so die oft postulierte Trennung von Grammatik und Lexikon aufheben.

2.1 PSYCHOLINGUISTIK

Als Muttersprachler wissen wir intuitiv, dass beispielsweise die Lautfolge [kni:p] ein deutsches Wort ergeben könnte, wohingegen [pmeh] nicht möglich ist. Zusätzlich wissen wir, ohne es explizit gelernt zu haben, dass es *Zähne putzen* heißt und nicht etwa *Zähne waschen* oder *bürsten*. Wir verfügen also über ein erstaunlich hohes Maß an Erfahrung und Wissen bezüglich der Sequenzen von kleineren und größeren Einheiten in unserer Sprache (Ellis 2002). Das deutet darauf hin, dass wir Elemente der Sprache nicht einzeln aus dem Gedächtnis abrufen, sondern diese auf allen linguistischen Ebenen, also bspw. Grapheme, Phoneme, Morpheme, Lexeme, Kollokationen und Phrasen zu Einheiten bzw. Chunks bündeln, die wir als Ganzes speichern. Der Begriff *Chunk* entstammt der Psychologie und geht zurück auf Miller (1956: 93), der damit eine Einheit in der Informationsverarbeitung bezeichnete:

We are dealing here with a process of organizing or grouping the input into familiar units or chunks, and a great deal of learning has gone into the formation of these familiar units.³

Die Kapazität unseres Kurzzeitgedächtnisses liegt bei ungefähr 7 ± 2 Chunks, wobei kleinere Chunks wiederum zu größeren Einheiten zusammengefasst werden können, sodass mehr Informationen verarbeitet werden können. Ellis (2003: 76) beschreibt anschaulich, dass wir uns auf Grund von *Chunking* zum Beispiel eine musterhafte Telefonnummer wie 0800-123777 oder auch eine Abfolge von Buchstaben wie AGREEMENTS besser merken können als beliebige Sequenzen (z.B. 4957-632518, HPLIUEBHZI). Analog prägen wir uns „*Ellis is wittering on about chunking again*“ leichter ein als „*again wittering on is*

³ „Hier geht es um einen Prozess, bei dem der Input in bekannte Einheiten oder Chunks geteilt bzw. dementsprechend strukturiert wird. Lernen besteht hauptsächlich aus der Bildung dieser Chunks.“ (eigene Übersetzung).

about Ellis chunking“ (ebd.; Hervorh. im Original). Demnach lässt sich festhalten: „We chunk chunks too [...]“ (ebd.). Chunking beschreibt also eine generelle kognitive Eigenschaft, bei der Sequenzen ganzheitlich abgerufen und nicht einzeln zusammengesetzt werden. Wray (2002: 9) bezeichnet diese Sequenzen sodann als formelhaft und definiert sie als:

[S]equence[s], continuous or discontinuous, of words or other meaning elements⁴, which [are], or appear[...] to be, prefabricated: that is stored and retrieved whole from the memory at the time of use, rather than being subject to generation or analysis by the language grammar.⁵

Gestützt wird diese Definition von Studien, die zeigen, dass formelhafte Sequenzen schneller verarbeitet werden. Dabei wird angenommen, dass eine hohe Frequenz im Input einer starken kognitiven Verankerung entspricht. Arnon und Snider (2010) zeigten bspw., dass erwachsene Muttersprachler häufig vorkommende Einheiten, die sogar eine Spanne von vier Wörtern umfassen, schneller als mögliche englische Ausdrücke identifizieren als weniger frequente Sequenzen. Dabei unterschieden sich die Vierworteinheiten als Ganzes in ihrer Frequenz (*don't have to worry* vs. *don't have to wait*), untergeordnete Einheiten verfügten jedoch über dieselben Frequenzen (*worry* ist genauso frequent wie *wait*, *have to worry* ist genauso frequent wie *have to wait* usw.), sodass der Effekt nicht auf einzelne frequente Wörter oder kleinere Wortfolgen zurückzuführen ist. Tremblay et al. (2011) verglichen die Geschwindigkeit mit der Sätze, die sogenannte *lexical bundles*, „frequently recurring strings of words that often span traditional syntactic boundaries“⁶ (ebd.: 569) (z.B. *in the middle of the*) enthielten sowie Sätze, die vergleichbare *non-lexical bundles* (z.B. *in the front of the*) enthielten, gelesen wurden und konnten zeigen, dass erstere Sätze schneller gelesen werden. Nicht nur Erwachsene, auch Kinder produzieren häufig vorkommende Sequenzen schneller als weniger frequente Einheiten (Bannard/Matthews 2008; s. 3.1).

⁴ Die vorliegende Arbeit beschränkt sich auf die Untersuchung von Mehrworteinheiten. Kleinere Elemente wie Phoneme oder Morpheme werden nicht berücksichtigt.

⁵ ‚(Dis-)kontinuierliche Sequenzen von Wörtern oder anderen bedeutungstragenden Elementen, die vorgefertigt sind bzw. vorgefertigt zu sein scheinen: das bedeutet, sie sind ganzheitlich gespeichert und werden im Sprachgebrauch ganzheitlich abgerufen, also weder zusammengesetzt noch analysiert.‘ (eigene Übersetzung).

⁶ ‚frequente, wiederkehrende Wortfolgen, die meist über syntaktische Grenzen hinausgehen‘ (eigene Übersetzung).

2.2 PHRASEOLOGIE

Traditionell machte es sich die Phraseologie zur Aufgabe, lexikalische Einheiten mit Hilfe von linguistischen Kriterien zu untersuchen und zu kategorisieren, wobei idiomatische Ausdrücke im Fokus der Analyse standen: „Idioms form the majority and may be regarded as the prototype of the phraseological unit⁷“ (Gläser 1998: 126; vgl. Granger/Paquot 2008: 28). Während sich dieser Ansatz dem Phänomen von außen (*top-down*) näherte, d.h. Phraseologismen wurden anhand bereits vorformulierter Kriterien bestimmten Kategorien zugeordnet, wählte Sinclair (1987) erstmals eine induktive (*bottom-up*) Herangehensweise, wobei er frequente Mehrworteinheiten aus großen Korpora extrahierte. Dabei wurde deutlich, dass restriktive Einheiten wie Idiome nur eine untergeordnete Rolle spielen, da sie relativ selten vorkommen, wohingegen weniger auffällige Muster, z.B. Kollokationen, den Sprachgebrauch in hohem Maße durchdringen (vgl. Granger/Paquot 2008: 29; vgl. Handwerker 2010: 249). Ebenso wenig wurden syntaktisch rekurrente Muster berücksichtigt, denn „[d]er Quasi-Lexemstatus und die Bestimmung als Wortverbindung blendet[e] den Aspekt syntaktischer Prägung weitgehend aus“ (Feilke 2007: 66).

Die bis hierhin verwendeten Bezeichnungen *Chunks*, *formelhafte Sequenzen* und *Phraseologismen* für den Untersuchungsgegenstand der vorliegenden Arbeit spiegeln die sehr heterogene Terminologie in der Forschung wider. So führen Wray und Perkins (2000: 3) über 40 Begriffe an, die genutzt werden, um Formelhaftigkeit in der Sprache zu beschreiben. Stumpf (2015: 21f.) konstatiert, dass die Begriffe Idiomatik, Phraseologie und Formelhaftigkeit folgendem Hyperonymie- bzw. Hyponymie-Verhältnis entsprechen: Formelhaftigkeit > Phraseologie > Idiomatik und so die Erweiterung des Gegenstandsbereichs der Phraseologie beschreiben:

Als ‚phraseologisch‘ werden also längst nicht mehr allein Einheiten mit idiomatischer Bedeutung angesehen, sondern auch solche Wendungen, die sich durch häufige Verwendung in fester Form auszeichnen. Der Bezeichnung ‚phraseologisch‘ kommt somit mittlerweile ein sehr breites Bedeutungsspektrum zu: Im engen (und klassischen) Sinne wird ‚phraseologisch‘ gleichgesetzt mit ‚idiomatisch‘, im weiten Sinne heißt ‚phraseologisch‘ (lediglich) ‚formelhaft‘.

(Stein 1994: 153)

Neben den terminologischen Wirrungen erscheint es nicht leicht, Merkmale zu finden, die eine klare Abgrenzung von formelhaften Wendungen zu „freien“ Sequenzen ermöglichen. „Frei“ steht hier in Anführungszeichen, da natürlich auch vermeintlich freie Sequenzen

⁷ ‚Idiome bilden die Mehrheit und können dementsprechend als Prototyp phraseologischer Einheiten gelten.‘ (eigene Übersetzung).

nicht beliebig zusammengesetzt werden, sondern Regeln der Syntax und Semantik folgen (vgl. Burger 2015: 19). Prototypisch werden die drei Eigenschaften Polylexikalität, Festigkeit und Idiomatizität zur Bestimmung von Phrasemen aufgeführt (vgl. ebd.: 14).

Polylexikalität meint, dass es sich bei den zu untersuchenden Einheiten um Mehrwortverbindungen und nicht um einzelne Wörter handelt, was zunächst recht unproblematisch scheint (vgl. ebd.: 15). Schwierigkeiten bei der Definition bereiten hierbei jedoch bspw. formelhafte Einwortäußerungen, die dieselben pragmatischen Funktionen wie äquivalente Mehrworteinheiten ausüben (*guten Tag* vs. *hallo*, *vielen Dank* vs. *danke*) (vgl. Stein 1995: 27). Als ebenfalls problematisch erweisen sich idiomatische Komposita wie *Angstphase* oder auch orthografische Konventionen (*auf Grund* vs. *aufgrund*) (vgl. Stumpf 2015: 26f.). Als Obergrenze von Phrasemen gilt der Satz; es scheint jedoch, dass Sprecher zumindest den Aufbau häufig gebrauchter Texte verinnerlicht haben, also die Struktur von Äußerungen, die über die Satzgrenze hinausgehen (vgl. Heinemann/Heinemann 2002: 135-140; vgl. Filatkina 2007: 141).

Das Merkmal Festigkeit besagt, dass Phraseme in der aktuellen Sprachverwendung „gebräuchlich“ sein müssen, wobei die Intuition eines Sprachwissenschaftlers nicht ausreicht, um dies zu belegen. So werden u.a. korpuslinguistische Analysen zur Frequenz und Kookkurrenz von Phrasemen immer wichtiger (s. 5.) (vgl. Burger 2015: 16). Nach Burger (vgl. ebd.: 16-26) kann Festigkeit in psycholinguistische, strukturelle und pragmatische Festigkeit unterteilt werden. Psycholinguistische Festigkeit greift das Konzept auf, dass vorgefertigte Sequenzen holistisch als Chunks gespeichert werden, was einer Überbelastung des Arbeitsgedächtnisses entgegenwirkt und so ein reibungsloses Sprechen ermöglicht. Strukturelle Festigkeit bedeutet, dass die Formseite einer musterhaften Sequenz gar nicht oder nur eingeschränkt verändert werden kann. Dies zeigt sich zum Beispiel bei Irregularitäten in der Morphologie und/oder Syntax, wie bei dem unflektierten attributiven Adjektiv in der Wendung *auf gut Glück*. Auch Restriktionen morphosyntaktischer oder lexikalisch-semantischer Art spielen hier eine Rolle (vgl. z.B. Schmitt/Carter 2004: 7), wobei zu beachten ist, dass eine strikte strukturelle Festigkeit eher die Ausnahme darstellt. Die meisten phraseologischen Einheiten lassen ein mehr oder minder breites Maß an Variation zu (vgl. Burger 2015: 19-24). Insbesondere Wortverbindungen, die auf Grund ihrer kommunikativen Funktion, d.h. ihrer pragmatischen Festigkeit einen formelhaften Charakter zugewiesen bekommen, sind strukturell als auch semantisch meist sehr heterogen. Sie werden genutzt, um „mit ihrer Hilfe rekurrente kommunikative Aufgaben bzw. kommunikative Rou-

tinehandlungen bewältigen ([zu] können)“ (Stein 2007: 226). Mit Stein (vgl. ebd.: 226f.) kann zwischen situationsgebundenen Einheiten, deren Verwendung an eine Kommunikationssituation geknüpft ist (z.B. *Guten Tag*) und die meist soziale Funktionen erfüllen und nicht-situationsgebundenen Einheiten (z.B. *wie eingangs erläutert*), die meist text- und interaktionsorganisierende Funktionen erfüllen, unterschieden werden.

Idiomatisch ist eine Wortsequenz dann, wenn die phraseologische Bedeutung losgelöst von der Bedeutung der beteiligten Elemente steht. So hat die Wendung *jmdm. einen Korb geben* im Sinne von *jmdn. zurückweisen* nichts mit der wörtlichen Bedeutung *jmdm. einen Korb überreichen* zu tun und gilt somit als idiomatisch. Nicht-idiomatisch hingegen ist der Ausdruck *die Zähne putzen*, da die Gesamtbedeutung aus den Einzelbedeutungen der enthaltenen Elemente ableitbar ist. Nach diesem Muster ist der Phraseologismus *blinder Passagier* teil-idiomatisch, da die Komponente *Passagier* ihre wörtliche Bedeutung beibehält. Stein (1995: 31) ergänzt, dass die wörtliche Bedeutung auch „zugunsten einer semantischen Reduzierung oder auch einer semantischen Leere, an deren Stelle eine oder mehrere kommunikative Funktionen treten“ aufgegeben werden kann, wie bspw. bei den situationsunabhängigen Wendungen *und so weiter*, *ich würde sagen* (vgl. Stumpf 2015: 37).

Gries (2008a: 4; Hervorh. im Original) beschreibt ebenfalls und unter Berücksichtigung korpuslinguistischer Analyseverfahren, zu welchen Parametern gelungene Definitionsversuche Stellung beziehen müssen:

- i. the *nature* of the elements involved in a phraseologism;
- ii. the *number* of elements involved in a phraseologism;
- iii. the *number of times* an expression must be observed before it counts as a phraseologism;
- iv. the permissible *distance* between the elements involved in a phraseologism;
- v. the degree of *lexical and syntactic flexibility* of the elements involved;
- vi. the role that *semantic unity* and *semantic non-compositionality / non-predictability* play in the definition.⁸

Dabei versteht Gries (vgl. ebd.: 5) Phraseologismus zunächst allgemein als das gemeinsame Auftreten eines lexikalischen Elements mit einem oder mehreren zusätzlichen linguistischen Elementen. Ein zusätzliches Element kann dabei ebenfalls einer Wortform entsprechen, wie bei *strong tea*; es kann sich jedoch auch um ein grammatisches Muster handeln

⁸ ‚(i) Welcher Art sind die Bestandteile eines Phraseologismus?, (ii) Wie viele Elemente enthält der Phraseologismus?, (iii) Wie oft muss ein Ausdruck vorkommen, um als Phraseologismus zu gelten? (iv) Wie weit dürfen Elemente eines Phraseologismus voneinander entfernt stehen?, (v) Welche Rolle spielen lexikalische und syntaktische Flexibilität?, (vi) Welche Rolle spielen semantische Einheit und Idiomatizität?‘ (eigene Übersetzung).

(bspw. wird das englische Verb *hem* meist im Passiv gebraucht – *to be hemmed in* ‚eingeklemmt sein‘). Als wichtigstes Kriterium müssen die Elemente eines Phraseologismus laut Gries (vgl. ebd.: 6f.) zusammen eine semantische Einheit bilden: dies zeigt sich oftmals durch Idiomatizität, meist existiert auch ein Quasi-Synonym (*jmdm. unter die Arme greifen: jmdm. helfen; jmdm. um den Hals fallen: jmdn. umarmen*), wobei Idiomatizität keine zwingende Eigenschaft von Phraseologismen ist:

the elements of a phraseologism [...] are usually assumed to function as a semantic unit, i.e. to have a sense just like a single morpheme or word. [...] For something to count as a phraseologism, I will require semantic unity, but not non-compositional semantics[.]⁹

(ebd.: 6)

Zusammenfassend wird Phraseologismus nach Gries (vgl. ebd.) verstanden als das gemeinsame Auftreten einer Wortform und eines oder mehrerer zusätzlicher sprachlicher Elemente unterschiedlicher Art, die als semantische Einheit fungieren und deren Frequenz die zu erwartende Frequenz übersteigt.

Das Kriterium der Frequenz ist nicht unproblematisch, da die Gefahr besteht, dass „niedrig frequente, aber hoch idiomatische Wendungen [...] nicht aufgedeckt, im Gegensatz dazu jedoch hochfrequente Verbindungen – jedweder Art – auf diese Weise als phraseologisch klassifiziert [werden]“ (Stumpf 2015: 40). Während Frequenzlisten rein quantitativ das Vorkommen von Mehrworteinheiten zeigen, berechnen Kookkurrenzmaße Textwörter, die

statistisch signifikant[...] [m]iteinandervorkommen [...]. Usuelle Kookkurrenzen sind in erster Instanz binäre Relationen zwischen autosemantischen Wortschatzelementen (Kollokationen). Dazu gehören auch all jene Wortverbindungen, die einen Mehrwortstatus aufweisen, also selbst als lexikalisch-semantische, grammatische und/oder pragmatische Einheiten anzusehen sind (z.B. Idiome, kommunikative Formeln, Funktionsverbgefüge usw.).

(Steyer 2003: 33)

Kookkurrenzanalysen erfassen also die gemeinsame Auftretensfrequenz von sprachlichen Einheiten und können so dabei helfen, Mehrworteinheiten zu extrahieren, deren einzelne Bestandteil insgesamt nicht unbedingt häufig, aber wenn, dann sehr oft miteinander vorkommen (s. 5.2.2).

Stumpf (2015: 49) bricht die Charakteristika formelhafter Sprache auf das „verbindende[...] Merkmal der Reproduzierbarkeit“ herunter (vgl. Stein 1995: 43). Steyer (2013: 28) betont ebenfalls das wiederholte Vorkommen „verwandter sprachlicher Strukturen in ver-

⁹ ‚Generell müssen die Elemente eines Phraseologismus zusammen eine semantische Einheit bilden, d.h. sie verfügen über eine Bedeutung ähnlich einem einzelnen Morphem oder Wort. Unter Phraseologismen verstehe ich Wortfolgen, die eine semantische Einheit bilden, sie müssen jedoch nicht zwingend idiomatisch sein.‘ (eigene Übersetzung).

gleichbaren Kontexten“ als grundsätzliche Bedingung für das Entstehen formelhafter Sequenzen, die sie *Usuelle Wortverbindungen* (UWV) nennt. Dabei fungieren Usuelle Wortverbindungen als autonome kommunikative Einheiten, d.h. sie sind

auch ohne weitere sprachliche Einheiten als komplexe[...] sprachliche[...] Ausdr[ü]ck[e] versteh- und einsetzbar [und haben] eine übersummativ (aber nicht unbedingt idiomatische) Qualität [...], die es ermöglicht, diese Entitäten als Ganzes abzurufen und im Kommunikationsprozess einzusetzen[.]

(ebd.: 24)

Darüber hinaus müssen sie nicht zwingend formale Irregularitäten aufweisen, auch grammatikalisch reguläre Einheiten können durch massenhaftes Auftreten zu festen Einheiten werden (vgl. ebd.: 28).

Steyer (2013: 24ff.) unterscheidet zwischen *Wortverbindungen*, deren lexikalische Bestandteile in Stellung, Schreibung und Flexion variieren können, ansonsten aber fest sind (z.B. *Geld abheben*, *Augen zu und durch*) und *Wortbindungsmustern*, die neben fixen lexikalischen Komponenten auch Leerstellen (slots) enthalten und denen eine morphosyntaktische Struktur, das ‚Strukturmuster‘ zugrunde liegt. Die gesamte Einheit wird als ‚Schema‘ bezeichnet (vgl. ebd.: 25; vgl. Tomasello 2006: 28ff.). (1) verdeutlicht, wie ein Schema durch unterschiedliche Strukturmuster, deren Leerstellen wiederum unterschiedlich lexikalisch gefüllt werden können, realisiert werden kann.

(1) Schema: X *davonkommen*

Strukturmuster 1:	ADJ <i>davonkommen</i> ADJ-Füller: <i>billig/ungeschoren/glimpflich</i>
Strukturmuster 2:	ADV-G <i>davonkommen</i> ADV-G-Füller: <i>einigermaßen/noch mal/gerade noch</i>
Strukturmuster 3:	PRÄPG <i>davonkommen</i> PRÄPG-Füller ¹⁰ : <i>mit einem blauen Auge/mit dem Schrecken/mit heiler Haut/ohne Strafe</i>

(Steyer 2013: 27)

Dabei spielen die gemeinsamen semantischen bzw. funktionalen Merkmale der Füller eine wichtigere Rolle als die morphosyntaktischen Kategorien, die ihnen gemein sind, da sie so die übersummativ Bedeutung des Schemas schaffen (vgl. ebd.).

Tabelle 1 zeigt den Versuch einer Typologie formelhafter Sequenzen, die semantische, strukturelle und pragmatische Kriterien berücksichtigt, wobei die einzelnen Klassen nicht eindeutig voneinander abzugrenzen sind und, im Hinblick auf die Analyse in 6., nur Phra-

¹⁰ Die Abkürzung mit Bindestrich ADV-G bedeutet, dass sowohl Adverbien als auch Adverbgruppen vorkommen können, wohingegen bei PRÄPG nur Präpositionalgruppen auftreten (vgl. Steyer 2013: 27).

seme unterhalb der Satzgrenze berücksichtigt wurden (vgl. Stumpf 2015: 47; vgl. Wotjak/Heine 2007: 42).

Klasse	Definition/Merkmale	Beispiele	Kriterium
Idiome	Bedeutung des komplexen Ausdrucks ist nicht gleich Summe der Einzelbedeutungen der jeweiligen Lexeme	<i>jmdm. reinen Wein einschenken, jmdn. an der Nase herumführen</i>	semantisch
Teil-Idiome	einzelne Komponenten weisen eine idiomatische, andere eine wörtliche Bedeutung auf	<i>einen Streit vom Zaun brechen, blinder Passagier</i>	semantisch
Kollokationen	Wendungen, die nicht oder nur schwach idiomatisch sind (häufige Struktur: N+V)	<i>sich die Zähne putzen, den Tisch decken</i>	semantisch/(strukturell)
Paarformeln/Zwillingsformeln	paarige Verbindungen zweier Wörter der gleichen Wortart mit <i>und</i> , einer anderen Konjunktion oder einer Präposition	<i>klipp und klar, weder Fisch noch Fleisch</i>	strukturell
komparative Phraseme	Vergleich, bestehend aus drei Hauptelementen: Vergleichsempfänger (jmd.), tertium comparationis (flink), Vergleichsspender (Wiesel)	<i>fressen wie ein Scheunendrescher, schlau wie ein Fuchs sein</i>	strukturell
Modellbildungen/Phraseoschablonen	Strukturschemata mit Leerstellencharakter, denen eine konstante semantische Interpretation zugeordnet ist	<i>von Stadt zu Stadt (von X zu X), typisch Mann (typisch X)</i>	strukturell
gesprächsspezifische Formeln	situationsunabhängige kommunikative Wendungen (keine selbstständigen Äußerungseinheiten)	<i>ich meine, wie gesagt</i>	pragmatisch
Routineformeln	situationsabhängige kommunikative Wendungen (selbstständige Äußerungseinheiten)	<i>gern geschehen, guten Tag</i>	pragmatisch
Funktionsverbgefüge	Zusammensetzungen aus einem nominalisierten Verb und semantisch „leeren“ Verben, mit denen die Aktionsart differenziert werden kann	<i>zur Entscheidung kommen/bringen/stellen/stehe, jmdm. den Vorzug geben</i>	strukturell/semantisch/lexikalisch
strukturelle Phraseme	besitzen die Funktion, (grammatische) Relationen herzustellen	<i>in Bezug auf, sowohl [...] als auch [...]</i>	strukturell, grammatisch

Tabelle 1. Typologie formelhafter Wendungen.

(leicht gekürzt, Stumpf 2015: 47ff.)

In Tabelle 1 werden unter anderem situationsunabhängige, kommunikative Wendungen berücksichtigt, die keine autonomen Äußerungseinheiten darstellen, sondern meist Inhaltliches, Komplexes einbetten. Diese widersprechen Steyers (vgl. 2013: 24) Definition von Usuellen Wortverbindungen, erfüllen jedoch das zentrale Kriterium der Reproduzierbarkeit, d.h. sie erfüllen wiederholt dieselben Funktionen bspw. in einem Text.

Zusammenfassend spricht also die Rekurrenz bzw. Reproduzierbarkeit für den Status von Mehrworteinheiten als einheitenbildende Chunks, die eine graduelle Ausprägung von lexikalischer Spezifiziertheit auf der einen und zunehmender Schematisierung auf der anderen Seite aufweisen und denen semantische, strukturelle und pragmatische Kriterien zugeordnet werden können (vgl. Wotjak/Heine 2007: 41). Diese Definition erscheint insofern als vielversprechend, als sie sich größtenteils mit neueren Ansätzen zur Beschreibung von Sprache und Theorien des Erst- und Zweitspracherwerbs deckt (s. 2.3, 3.). Diese gehen

von Form-Bedeutungspaaren als Basiseinheiten der Sprache aus, sodass die traditionelle Trennung von Lexikon und Grammatik annulliert wird und Phraseme nicht länger nur als Randerscheinungen betrachtet werden.

2.3 SPRACHTHEORETISCHE VERORTUNG

Sollte der sprachliche Normalfall nicht das Okkasionelle, sondern das Geprägte und Vorgeprägte sein, müssten Theorien und Methoden der Linguistik darauf umgestellt werden, dass die ad-hoc-Bildungen mehr oder weniger kreative Realisierungen idiomatisch geprägter und vorgeprägter Modelle darstellen, dass also die grammatischen Regeln auf der Basis dieser Modelle funktionieren.

(Ágel 2004: 67)

Der linguistische Strukturalismus, der maßgeblich die Sprachwissenschaft des 20. Jh. prägte, unterscheidet zur Beschreibung und Erklärung von Sprache zwischen Sprachgebrauch einerseits und Sprache als abstraktem linguistischem System andererseits. Saussure (vgl. 1916: 33f., 38) führt dafür das Begriffspaar *langue* (Sprachsystem) und *parole* (konkrete, individuelle Realisierungen von Sprache) an, Chomsky (vgl. 1965: 10) greift diese Dichotomie von Sprache und Sprechen in der Generativen Grammatik¹¹ mit den entsprechenden Bezeichnungen *Kompetenz* (sprachliches, unbewusstes Wissen eines idealisierten Sprecher-Hörers) und *Performanz* (Sprachgebrauch) auf. Lediglich die tieferliegende Struktur von Sprache wird zum Untersuchungsgegenstand des Strukturalismus. Ziel war es unter anderem, die Kompetenz von Muttersprachlern zu beschreiben, die sich vor allem in der Fähigkeit manifestiert, grammatisch wohlgeformte von ungrammatischen Sätzen unterscheiden zu können (vgl. Chomsky 2002: 13). Der Generativen Grammatik zufolge nutzen Muttersprachler dafür ein System von algebraischen Regeln, das bedeutungsvolle Elemente zu grammatischen Sätzen kombiniert, wobei die Bedeutungen bei der Zusammensetzung keine Rolle spielen. Diese sind separat im Lexikon enthalten, das darüber hinaus die phonologische Form sowie die syntaktische Kategorie von lexikalischen Einheiten enthält (vgl. Chomsky 1995: 30; vgl. Gries 2008a: 10; vgl. Pawley 1986: 59). Vereinfacht ausgedrückt: bedeutungstragende Wörter werden mit Hilfe von bedeutungslosen Grammatikregeln zusammengefügt, Eigenschaften von Wörtern sowie Ausdrücke, die sich nicht aus Regeln ableiten lassen, sind im Lexikon gespeichert. Dieses Modell betrachtet das Sprachvermögen außerdem als autonomes Subsystem, das unabhängig von anderen kognitiven Modulen, wie z.B. der Wahrnehmung oder dem Gedächtnis, arbeitet (vgl. Chomsky 1995:

¹¹ Grammatik meint hier und im Folgenden nicht etwa ein Regelwerk zur Beschreibung von Sprache wie bspw. die Duden-Grammatik, vielmehr bezeichnet Grammatik eine Theorie, die beschreiben und erklären muss, wie Sprache repräsentiert und erworben wird (vgl. Ziem/Lasch 2013: 8; vgl. Chomsky 1981: 35f.).

167).

Die einseitige Beschreibung von Kompetenz ohne Berücksichtigung des tatsächlichen Sprachgebrauchs, die fehlende Auseinandersetzung mit Irregularitäten, wie formelhafte Sequenzen sie häufig aufweisen, sowie der modulare Ansatz wurden unter anderem von Ellis (2012: 194; vgl. auch Tomasello 2003: 5) kritisiert:

Grammar was modularized, encapsulated, and divorced from performance, lexis, social usage, and the rest of cognition. Patterns, structures, constructions, formulaic language, phraseology, all were peripheral.¹²

Es wird gefordert, Sprache in Übereinstimmung mit generellen kognitiven Kenntnissen und Fähigkeiten des Menschen zu beschreiben: „to make one’s account of human language accord with what is generally known about the mind and the brain, from other disciplines as well as our own“ (Lakoff 1990: 40; vgl. Robinson/Ellis 2008: 3). Kognitive Linguisten gehen davon aus, dass es kein unabhängiges sprachspezifisches Kenntnissystem gibt, d.h., dass die Organisation und der Erwerb sprachlichen Wissens der Organisation und dem Erwerb nichtsprachlichen Wissens ähneln: „knowledge of language is knowledge“ (Goldberg 1995: 5; vgl. Croft/Cruse 2004: 2; vgl. Langacker 2013: 8)¹³.

Darüber hinaus wird postuliert, Sprache umfassend und vollständig zu beschreiben und dabei nicht zwischen zentralen Elementen einer Grammatik und Ausnahmen zu unterscheiden, denn Regularität ist nicht die Norm (vgl. Ziem/Lasch 2013: 18). Ebenso wenig stehen sich Lexikon und Grammatik isoliert gegenüber als ein Pool von bedeutungstragenden Lexemen, die mit Hilfe von Grammatikregeln zusammengefügt werden. Auch grammatische Strukturen verfügen über Bedeutung. So bemerkte bspw. Fillmore (1968: 48), dass folgende Sätze eine andere Bedeutung haben, obwohl die lexikalischen Elemente dieselben sind:

- (2) Bees are swarming in the garden.
- (3) The garden is swarming with bees.

¹² ‚Grammatik wurde modularisiert, zusammengefasst und so losgelöst vom (sozialen) Sprachgebrauch, der Lexik und anderen kognitiven menschlichen Eigenschaften untersucht. Muster, Strukturen, Konstruktionen, formelhafte Sprache, Phraseologie, all das wurde in die Peripherie abgeschoben.‘ (eigene Übersetzung).

¹³ Dafür spricht bspw., dass beim Textverstehen sprachliches Wissen mit nicht-sprachlichem Wissen interagiert, wie Bransford et al. (1972) nachweisen konnten. Andererseits sprechen Ergebnisse der Aphasieforschung (mit Aphasien werden Sprachstörungen bezeichnet, die meist auf Grund von Durchblutungsstörungen im Gehirn entstehen) für den Status von Sprache als eigenständigem mentalem Modul, denn „[b]ei aphasischen Störungen sind die allgemeinen kognitiven Fähigkeiten meistens nicht beeinträchtigt“ (Schwarz 2008: 92). Auf die unterschiedlichen Ansätze kann an dieser Stelle nicht weiter eingegangen werden. Ich schließe mich jedoch Schwarz (vgl. ebd.: 58) an, die den fehlenden wissenschaftlichen Austausch beider Ansätze kritisiert und für eine verstärkte Zusammenarbeit plädiert.

Bei (3) wird davon ausgegangen, dass der ganze Garten voll mit Bienen ist, wohingegen die Bienen bei (2) auch nur einen Teil des Gartens bevölkern könnten. Goldberg (vgl. 2013: 18f.) zeigt, wie die Ditransitiv-Konstruktion in (4) generell mit der Bedeutung „Transfer, Weitergabe“ verknüpft ist.

(4) She gave him a book.

So interpretierten 60 Prozent der Teilnehmenden das unsinnige Wort *moop* in (5) als *give* ‚geben‘, d.h. die Funktion „Transfer“ ist der ditransitiven Struktur und nicht etwa nur dem Verb *give* inhärent. Die restlichen 40 Prozent gaben Antworten, die nichtsdestoweniger die Bedeutung von Transfer, wortwörtlich oder im übertragenen Sinne enthielten, wie *tell* ‚berichten, erzählen‘ (Ahrens 1995, zitiert nach Goldberg 2013: 19).

(5) She mooped him something.

2.3.1 KONSTRUKTIONSGRAMMATIK

Konstruktionsgrammatische Ansätze geben konsequenterweise die Trennung von Lexikon und Syntax auf, das sprachliche System besteht nunmehr allein aus symbolischen Strukturen (‚symbolic structures in Cognitive Grammar‘, vgl. Langacker 2013) bzw. Konstruktionen (‚constructions‘, vgl. Bybee 2008¹⁴; vgl. Goldberg 2006), das heißt, aus konventionalisierten Form-Bedeutungspaarungen, die nicht nur Morpheme und Wörter, sondern auch abstrakte grammatikalische Muster enthalten.¹⁵

All levels of grammatical analysis involve constructions: learned pairings of form with semantic or discourse function [...] Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency.¹⁶ (Goldberg 2006: 5)

¹⁴ Bybee (2008: 220) verwendet u.a. auch den Begriff *Chunk*: „In Usage-Based Theory, constructions are chunks [...] with movable parts. They are established through practice and processed as single units“ „In gebrauchsbasierten Theorien sind Konstruktionen Chunks mit flexiblen Bestandteilen. Sie entstehen im Sprachgebrauch und werden ganzheitlich verarbeitet“ (eigene Übersetzung).

¹⁵ Unter „Konstruktionsgrammatik“ werden eine Vielzahl von Ansätzen gefasst, die folgende Ansichten teilen: a) Grammatische Konstruktionen gelten als Form-Bedeutungspaare; b) sprachliche Zeichen und grammatische Strukturen sind einheitlich als Konstruktionen repräsentiert; c) Grammatik gilt als strukturiertes Inventar von Konstruktionen (vgl. Ziem/Lasch 2013: 36; vgl. Fischer/Stefanowitsch 2006: 4f.); s. auch Arbeitskreis „Konstruktionsgrammatik in Deutschland, Österreich und der Schweiz“ <<https://konstruktionsgrammatik.wordpress.com>> (01.11.16).

¹⁶ „Alle Ebenen der grammatischen Analyse umfassen Konstruktionen: erlernte Paare aus Form- und Bedeutungskomponenten. Eine sprachliche Struktur gilt nur dann als Konstruktion, insofern irgendein form- oder bedeutungsseitiger Aspekt sich nicht auf der Grundlage der beteiligten Elemente der Struktur oder von bereits etablierten Konstruktionen vorhersagen lässt. Ebenso gelten Muster als Konstruktionen, die, obwohl vorhersagbar, oft genug vorkommen.“ (eigene Übersetzung).

Mit anderen Worten, Konstruktionen beschreiben das gesamte Syntax-Lexikon-Kontinuum, sie reichen von Morphemen und Wörtern bis hin zu schematischen Strukturen, Wortarten und grammatischen Relationen (vgl. Ziem/Lasch 2013: 18). Die Formseite umfasst syntaktische sowie morphologische und phonetische Eigenschaften der Konstruktion und die Bedeutungsseite enthält neben semantischen auch pragmatische und diskursfunktionale Aspekte. So werden Syntax, Semantik und Pragmatik nicht auf unterschiedlichen sprachlichen Ebenen getrennt voneinander beschrieben (vgl. Welke 2013: 21).

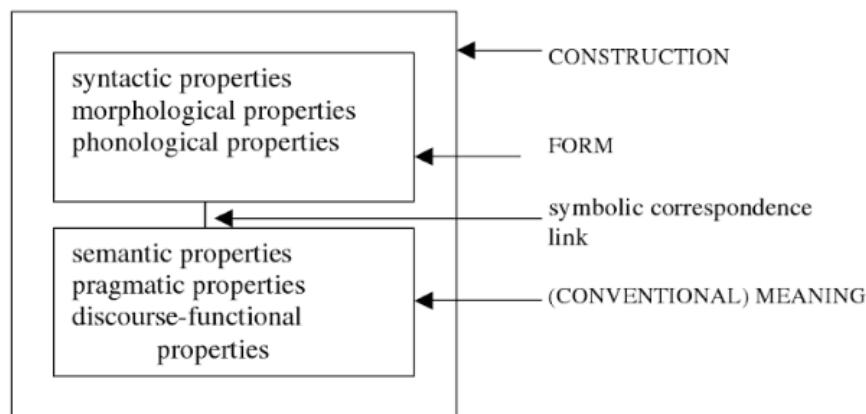


Abbildung 1. Konstruktionen als Form-Bedeutungspaare.

(Croft 2001: 18)

In früheren Arbeiten Goldbergs (vgl. 1995: 4) galten sprachliche Muster nur dann als Konstruktionen, wenn sie das Kriterium Unvorhersagbarkeit erfüllten. Formal sind z.B. die Komposita *Apfelgetränk* und *Beerengeränk* nicht vorhersagbar, da *Apfel* im Singular steht während *Beeren* die Pluralform bzw. eine Verbindung der Kompositionsglieder mit Fugen-*n* aufweist. Die Bedeutung lässt sich jedoch aus den einzelnen Elementen ableiten (vgl. Stefanowitsch 2011a: 185). Umgekehrt wird die Form *Trinker* regelhaft aus „*trink-/*,Flüssigkeit in den Mund nehmen und schlucken““ und „*[V_x-er]*,jemand, der regelmäßig X tut““ gebildet, der kursiv markierte Bedeutungsaspekt „jemand, der regelmäßig *und zwanghaft alkoholhaltige Flüssigkeit* in den Mund nimmt und schluckt““ (ebd.; Hervorh. im Original) bleibt jedoch unvorhersehbar. Die Konstruktion „*jmd. unter den Tisch trinken*“ ist sowohl semantisch als auch formal nicht vorhersagbar. Auf der Bedeutungsseite ist die Dimension „alkoholhaltige Getränke“ nicht aus den einzelnen Wörtern ableitbar, auf der Formseite ist die Konstruktion insofern unvorhersagbar, als dass sie eingeschränkt ist, „**jmd. unter den Stuhl trinken*“ ist bspw. nicht möglich (ebd.; Hervorh. im Original). Die „what’s-X-doing-Y“-Konstruktion („What’s the fly doing in my soup?“) (Kay/Fillmore 1999; vgl. Ziem/Lasch 2013: 13) veranschaulicht, wie pragmatische Eigenschaften mit

einer sprachlichen Form assoziiert sein können, da die Frage keiner Frage im eigentlichen Sinn entspricht, vielmehr soll die Überraschung in der Situation zum Ausdruck gebracht werden.

Stefanowitsch (vgl. 2011a: 184f.) plädiert dafür, zwischen *Konstruktionen*, d.h. Strukturen, die syntaktisch und/oder semantisch nicht transparent sind und *Mustern*, d.h. regelhafte, kompositionelle Strukturen, die so frequent sind, dass sie ebenfalls ganzheitlich gespeichert werden, zu unterscheiden. Da in dieser Arbeit (innerhalb der Konstruktionsgrammatik) gebrauchsbasiert-kognitive (*usage-based*) Beschreibungsansätze von Sprache fokussiert werden, die davon ausgehen, dass eine hohe Gebrauchshäufigkeit zu kognitiver Verfestigung und zu ganzheitlicher Speicherung führt, wird kein Unterschied zwischen Konstruktionen und Mustern gemacht (Bybee/Hopper 2001, Tomasello 2003). Dies deckt sich mit Goldbergs (2006: 18; Hervorh. im Original) Prämisse „**it’s constructions all the way down**“. Boas (vgl. 2013: 235) ergänzt, dass die Art der Bedeutung, die Konstruktionen inhärent ist, je nach Abstraktionsgrad, variiert, d.h. bspw. Inhaltswörter verfügen über detaillierte Bedeutungen, wohingegen abstraktere Konstruktionen wie das Passiv nur vage Bedeutungen (z.B. im Falle des Passivs einen Perspektivwechsel) in sich tragen (vgl. auch Stefanowitsch 2011b: 18).

Festzuhalten bleibt, (i) dass mit Konstruktionen auch sprachliche Phänomene, die bislang als Idiosynkrasien behandelt wurden, in die Beschreibung einbezogen werden und (ii) dass Konstruktionen zusätzliche Bedeutungskomponenten, die sich nicht aus den Verbbedeutungen oder den Bedeutungen anderer sprachlicher Elemente eines Satzes ergeben, erklären können (vgl. Ziem/Lasch 2013: 17-26). Gerade in der Bereitstellung pragmatischer Informationen liegt ein wesentliches Potenzial der Konstruktionsgrammatik für DaF, da „[d]er Fokus [...] auf der Erfassung semantisch und pragmatisch angemessener Äußerungen [liegt], im Gegensatz zur Erzeugung potenziell wohlgeformter Sätze in anderen Modellen“ (Handwerker 2008: 59).

2.3.2 KONSTRUKTIKON

Gebrauchsbasierte Modelle gehen also davon aus, dass sich Konstruktionen als Form-Bedeutungspaare nicht nur zur Beschreibung von Sprache auf der Kompetenzebene eignen, sondern dass sie gleichzeitig Einheiten der Performanz sind, denn

sie wandeln sich einerseits unter den Bedingungen des individuellen Sprachgebrauchs und sie bilden andererseits das Sprachwissen, über das SprachbenutzerInnen verfügen müssen, um eine Sprache [...] verstehen und verwenden zu können[.]

(Ziem/Lasch 2013: 95)

Demnach stellt sich die Frage, wie unser Sprachwissen organisiert ist. Goldberg zufolge sind die sprachlichen Informationseinheiten bzw. Konstruktionen in einem hochgradig strukturierten Netzwerk, dem Konstruktikon, verankert, wo sie bezüglich ihrer Form- und Bedeutungsseite über Vererbungshierarchien und Beziehungen zueinander verbunden sind (vgl. Ziem/Lasch 2013: 97f.; vgl. Ziem 2014: 23). Bei formseitigen Vererbungsrelationen handelt es sich um Teil-Ganzes-Beziehungen. Dabei vererben abstraktere Konstruktionen, die höher anzusiedeln sind, ihre Eigenschaften auf niedriger gestellte Konstruktionen, wie in Abbildung 2 dargestellt.

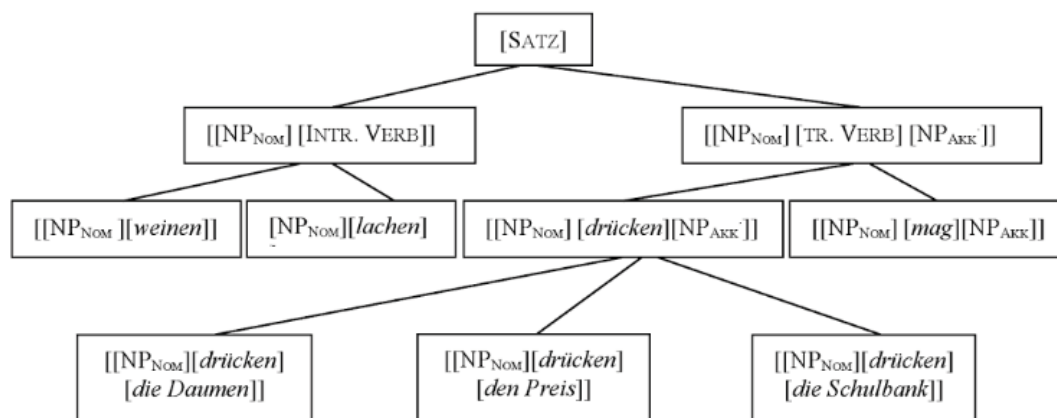


Abbildung 2. (Ausschnitt aus einem) Netzwerk formseitig verbundener Konstruktionen.

(Ziem/Lasch 2013: 97)

Es herrscht allerdings keine Einigkeit darüber,

bis zu welchem Abstraktionsgrad Konstruktionen sinnvollerweise als feste Einheiten im Sprachwissen postuliert werden können, ohne die zentrale Zielvorgabe aufzugeben, psychologisch realistische Beschreibungen vorzulegen.

(Ziem 2014: 21)

Während formseitig nur von Teil-Ganzes-Beziehungen ausgegangen wird, werden inhaltsseitig verschiedene Relationen postuliert wie Polysemie-Relationen – z.B. *Linnea gibt Henrike ein Stück Kuchen*. (X VERURSACHT, DASS Y Z ERHÄLT) und *Henrike erlaubt Peter*

eine Bootsfahrt. (X ERMÖGLICHT Y, Z ZU ERHALTEN) (Ziem/Lasch 2013: 99) –, Teil-Ganzes-Relationen – z.B. *Klaus malt.* und *Klaus malt den Tisch blau.* (ebd.: 100) –, Instanz- bzw. Beispiel-von-Relationen – z.B. $[[NP_{\text{NOM}}][treiben][NP_{\text{AKK}}][PP]]$ und *Georg treibt Klaus in den Wahnsinn.* (ebd.: 101) – und Relationen der metaphorischen Erweiterung – z.B. kann eine Resultativkonstruktion wie *Er küsst sie in Ekstase.* als metaphorische Erweiterung einer „Caused-Motion“-Konstruktion wie *Er wirft sie aufs Bett.* angesehen werden, insofern die Präpositionalphrase *aufs Bett* ein konkretes Ziel beschreibt, wohingegen *in Ekstase* eine innere Zustandsänderung bezeichnet¹⁷ (ebd.) – (vgl. Goldberg 1995: 74-81; vgl. Ziem 2014: 26)¹⁸.

Alternativ sei auf das FrameNet-Konstruktikon verwiesen, das einen ersten Versuch darstellt, Sprache unter Berücksichtigung der Konstruktionsgrammatik und der Framesemantik¹⁹ zu beschreiben und

das es erlaubt, Bedeutung(en) von Konstruktionen variierender Komplexität – von lexikalischen Einheiten bis zu komplexen grammatischen Konstruktionen – in ihrem Zusammenhang darzustellen [, wobei dies bis jetzt] erst ausschnitthaft und exemplarisch für die Gegenwartssprache Englisch erarbeitet [wurde.]

(Ziem 2014: 29; vgl. Fillmore/Baker 2015: 815f.)

Es bleibt festzuhalten, dass die Forschung bezüglich der Organisation des Konstruktikons noch am Anfang steht, Annahmen je nach Theorie variieren und viele weitere Korpusanalysen notwendig sind, um neue Erkenntnisse in Bezug auf den Aufbau des Konstruktikons zu gewinnen (vgl. Ziem/Lasch 2013: 102; vgl. Ziem 2014: 24).

Zusammenfassend soll nun auf das Verhältnis von Phraseologie und Konstruktionsgrammatik eingegangen werden. Gries (vgl. 2008a: 14f.) konstatiert, dass Kognitive Grammatik und Konstruktionsgrammatik im Gegensatz zur Generativen Grammatik sehr gut mit phraseologischen Ansätzen harmonieren, der einzige terminologische Unterschied darin bestehend, dass ein Phraseologismus wenigstens ein lexikalisches Element enthalten muss, während symbolische Strukturen bzw. Konstruktionen auch abstrakte, lexikalisch offene Muster zulassen (vgl. auch Dobrovol'skij 2011). Auch Steyer (vgl. 2013: 37) beleuchtet die

¹⁷ Die Metapher ist also „Change of State as Change of Location“ (Goldberg 1995: 99).

¹⁸ Diese Arten der Relation nach Goldberg wurden jedoch auch kritisiert (vgl. Boas 2003; vgl. Kay 2005).

¹⁹ Die Framesemantik untersucht, wie Menschen sprachliche Formen, die von Wörtern bis zu grammatischen Strukturen reichen, vor dem Hintergrund kognitiver Strukturen, d.h. Frames, die aktiviert werden, interpretieren und verstehen (Fillmore/Baker 2015). Beispiele für Frames sind u.a. das Wissen über die Abfolge und über beteiligte Personen bei einem Restaurantbesuch oder die Etappen und Prozesse im Leben eines Menschen (vgl. ebd.: 792). Das FrameNet Projekt (<https://framenet.icsi.berkeley.edu>), eine Datenbank für das Englische, befasste sich ausschließlich mit lexikalischen Elementen, das FrameNet-Konstruktikon stellt diesbezüglich also eine Erweiterung um grammatische Konstruktionen dar.

Überschneidungen von Usuellen Wortverbindungen und Konstruktionen und betont, dass konstruktionsgrammatische Arbeiten primär die Syntax fokussieren, während bei ihrem Ansatz vor allem die lexikalischen Komponenten bzw. deren Bedeutungen untersucht werden. Analog dazu lässt sich Stumpfs (2015: 22) Einordnung von Idiomatik, Phraseologie und Formelhaftigkeit um das Konzept der Konstruktion erweitern: „Der aus der Konstruktionsgrammatik stammende Begriff der Konstruktion (Form-Bedeutungspaar) kann als oberstes Hyperonym angesehen werden“.

3. CHUNKS IM ERST- UND ZWEITSPRACHERWERB

Im Folgenden soll beleuchtet werden, welche Bedeutung Chunks im Erst- und Zweitspracherwerb zuzukommen ist und welche Faktoren beim Erwerb von Konstruktionen eine Rolle spielen.

3.1 CHUNKS IM ERSTSPRACHERWERB

Dass Chunking eine zentrale Rolle im Erstspracherwerb spielt, scheint unumstritten. Kinder übernehmen Mehrwortsequenzen aus dem Input ohne diese zu analysieren (vgl. Bannard/Lieven 2012: 3; vgl. Wray 2002: 105). Lieven et al. (2009) untersuchten bspw. die Sprache von vier Zweijährigen mit Hilfe der sogenannten „Traceback“-Methode, wobei die Aufnahmen einer Zeitspanne (zwei Stunden) mit bereits attestierten Äußerungen verglichen werden (den 28 vorangegangenen Stunden) und fanden heraus, dass die Kinder 20 bis 40 Prozent der Äußerungen in derselben Form schon einmal produziert hatten, was eine ganzheitliche Speicherung der Sequenzen annehmen lässt. Bannard und Matthews (2008) extrahierten häufig vorkommende Chunks aus einem Korpus mit ‘an-das-Kind-gerichteter Sprache’ (*child-directed speech*) und kamen zu dem Ergebnis, dass Kinder im Alter von zwei und drei Jahren frequente Sequenzen (z.B. *a drink of milk*) schneller und eher fehlerfrei wiederholen als weniger frequente Chunks (z.B. *a drink of tea*).

Unter Rückbezug auf die Generative Grammatik (s. 2.3) soll nachfolgend kontrastiv Tomasello Modell (2003, 2006) einer gebrauchsbasierten Spracherwerbstheorie dargestellt werden, der annimmt, dass Kinder ausgehend von konkreten Chunks mit spezifischen Wörtern immer komplexere und abstraktere Konstruktionen lernen, indem sie Regularitäten aus dem Sprachgebrauch ableiten.

3.1.1 NATIVISTISCH VS. GEBRAUCHSBASIERT

Um die Frage, wie Kinder muttersprachliche Kompetenz erwerben, zu beantworten, muss z.B. eine Erklärung für die Tatsache gefunden werden, dass Kinder mehr über Sprache lernen, als die sprachlichen Daten, die sie beobachten, „hergeben“:

[...] the so-called *logical problem of language acquisition* or the problem of *the poverty of the stimulus*, in other words the mismatch between the input that children are exposed to and their ultimate attainment[.]²⁰

(White 2007: 37; Hervorh. im Original)

Chomsky (vgl. 1981: 35) geht davon aus, dass allen natürlichen Sprachen einige universelle Prinzipien zugrunde liegen. Dieses Sprachwissen muss nicht erlernt werden, es ist genetisch im Menschen als *Universalgrammatik* (UG) angelegt. Das erklärt z.B., warum Kinder und Erwachsene nie zuvor vernommene Sätze verstehen und bilden können (vgl. White 2007: 39). Die Annahme einer Universalgrammatik birgt jedoch einige ungelöste Probleme. Es stellen sich bspw. folgende Fragen (Tomasello 2003: 7):

Wie verknüpft ein Kind die abstrakten, angeborenen Strukturen mit den Besonderheiten der Zielsprache, die es lernt (‘the linking problem’)?

Wie erklärt sich die Entwicklung der Sprachfähigkeit von Kindern, wenn die UG immer dieselbe bleibt (‘the problem of continuity’)?

Diese Dilemmata entfallen, wenn man die zentrale Stellung des in nativistischen Erwerbstheorien wenig beachteten Sprachgebrauchs und dessen Einfluss auf die kognitive Repräsentation von Sprache anerkennt. Mit anderen Worten, kognitive Linguisten lehnen eine hypothetische UG ab und argumentieren, dass Kinder Konstruktionen, also Form-Bedeutungseinheiten, in ihrer situativen Gebundenheit aus dem Input ableiten, wobei diese nach und nach immer komplexer und schematischer werden (Robinson/Ellis 2008). Mit jedem neuen Input, den Lerner erfahren, erhalten sie weitere statistische Informationen:

All lexical items are primed for grammatical and collocational use, i.e. every time we encounter a lexical item it becomes loaded with the cumulative effects of those encounters, such that it is part of our knowledge of the word that it regularly co-occurs with particular other words or with specific grammatical functions[.]²¹

(Hoey 2004: 21)

²⁰ ‘das sogenannte logische Problem des Spracherwerbs bzw. das Poverty-of-the-Stimulus-Argument, m.a.W. das Missverhältnis zwischen dem, was Kinder im Input erfahren und dem, was sie schlussendlich können’ (eigene Übersetzung).

²¹ ‘alle lexikalischen Elemente werden auf grammatischen und kollokativen Gebrauch geprimed, d.h. jedesmal, wenn wir einem lexikalischen Element begegnen, erweitern wir unser Wissen zu diesem Element. So speichern wir als Teil des Wortwissens, dass es regelmäßig mit bestimmten anderen Wörtern oder bestimmten grammatischen Funktionen auftritt’ (eigene Übersetzung).

Vor diesem Hintergrund wird das Ziel, einen muttersprachlichen Endstatus oder die Kompetenz eines idealisierten Sprecher-Hörers zu beschreiben, hinfällig (vgl. Ellis 2003: 64). Die Sprachkompetenz eines Erwachsenen wird vielmehr als ein strukturiertes Inventar von Konstruktionen betrachtet, wobei Konstruktionen, die vielen anderen Konstruktionen ähneln eher im Kernbereich und weniger ähnliche eher am Rand angesiedelt sind. Alle Konstruktionen werden dabei mit den gleichen Lernprozessen²² erworben, die zudem komplexer und ausgefeilter sind, als zunächst angenommen²³ (vgl. Tomasello 2003: 3-6).

Tomasellos (vgl. ebd.: 325-328) Ansatz basiert also auf drei fundamentalen Annahmen: (i) Kinder lernen Sprache, um kommunikative Absichten verstehen und ausdrücken zu können (*intention-reading*), (ii) Konstruktionen und nicht einzelne Wörter oder Morpheme bilden dabei die Basiseinheiten im Spracherwerb und (iii) Kinder speichern konkrete Sequenzen, die sie im *Input* hören und in denen sie Muster identifizieren.

Bei Kindern sind die Konstruktionen noch nicht so sehr vernetzt wie bei Erwachsenen, d.h. sie entsprechen eher einzelnen „Inseln“ (vgl. Lieven/Tomasello 2008: 170). Anfangs werden konkrete Dinge, wie *cat* ‚Katze‘ und Holophrasen wie *Pick-up!* ‚Heb-auf‘ erlernt. In einem nächsten Schritt sind Kinder in der Lage Schemata zu benutzen, die aus mehreren Wörtern, meist aus einem invarianten Bestandteil und variablen Elementen bestehen, z.B. *Where's-the X?* ‚Wo-ist-der/die/das X?‘ oder *X gone* ‚X weg‘. Die dritte Entwicklungsstufe umfasst itemgestützte Konstruktionen, die bereits syntaktische Markierungen (bspw. Wortstellung) enthalten und die Perspektive eines Teilnehmenden in einem Szenario ausdrücken: *X hit Y* ‚X schlägt Y‘ oder *Y broken* ‚Y kaputt‘. An einem gewissen Punkt sind Kinder in der Lage, rein schematische Konstruktionen aus dem Sprachgebrauch abzuleiten, sowie Konstruktionen zueinander in Verbindung zu setzen²⁴ (vgl. Lieven/Tomasello 2008: 168-171; vgl. Tomasello 2006: 23f.; Diessel 2013).

²² - *intentional reading*, das Kleinkind begreift sich und den Erwachsenen als intentional handelnde Personen und kann kommunikative Absichten des Erwachsenen verstehen, und *pattern-finding*, das Erkennen von sprachlichen Mustern - (für einen Überblick kognitiver Prozesse s. auch Croft/Cruse 2004: 40-73)

²³ Vgl. Chomskys (1959) Kritik an behavioristischen Spracherwerbstheorien.

²⁴ Ob und wann Kinder verschiedene Form-Bedeutungseinheiten vergleichen und verknüpfen ist noch nicht geklärt (vgl. Lieven/Tomasello 2008: 171).

3.1.2 FAKTOREN IM GEBRAUCHSBASIERTEN ERSTSPRACHERWERB

In konstruktionsgrammatischen gebrauchsbasierten Sprachtheorien wird also davon ausgegangen, dass Kinder mit Hilfe allgemeiner kognitiver Prozesse wie Kategorisierung, Assoziierung und Automatisierung (vgl. Bybee 2013: 49f.) zunächst wenig variable Konstruktionen mit bestimmten Wörtern lernen, die nach und nach immer komplexer werden und aus denen Kinder abstrakte Regularitäten deduzieren. Dabei beeinflussen Frequenz und Form sowie das Zusammenwirken dieser Faktoren das assoziative Erlernen von Konstruktionen. Dies soll im nächsten Abschnitt näher erläutert werden.

3.1.2.1 FREQUENZ

Unsere Sprachverarbeitung spiegelt die Nutzungsfrequenz von sprachlichen Einheiten auf allen linguistischen Ebenen wider (Ellis 2002). Auch unsere Wahrnehmung beruht auf Wahrscheinlichkeiten, die wir dem Sprachgebrauch entnommen haben bzw. entnehmen. Wir wissen implizit, welche Konstruktionen wahrscheinlicher als andere sind. Da *one* ‚eins‘ bspw. öfter als *won* ‚gewinnen‘ vorkommt, tendieren wir zu ersterer Interpretation, wenn wir die Lautfolge [wʌn] hören (es sei denn, wir verfügen über zusätzliche Informationen). Ferner wissen wir, welche Bedeutung in einem bestimmten Kontext am wahrscheinlichsten ist, sodass wir bei *Alice in ...* automatisch an [wʌn]*derland* und nicht an die Ziffer *one* denken (vgl. Ellis 2008: 374). Ellis (2013: 369) betont: „These frequency effects are thus compelling evidence for usage-based models of language acquisition, which emphasize the role of input“²⁵.

Um weiter zu verdeutlichen, wie sich die Frequenz auf den Spracherwerb auswirkt, werden zwei Arten von Frequenz unterschieden (vgl. Bybee 2008: 218; vgl. Lieven/Tomasello 2008: 174): *Token-Frequenz* bezeichnet die Anzahl einer bestimmten Wortform oder Sequenz im Input, d.h. wie oft diese bestimmte Form auftritt²⁶; *Type-Frequenz* zeigt an, wie viele unterschiedliche Elemente einem bestimmten Muster bzw. einer Leerstelle zugeordnet werden können. Eine hohe Token-Frequenz führt zu Automatisierung; Sprachliche Formen werden schneller verarbeitet und ganzheitlich gespeichert. So lernen Kinder den Ausdruck *What's that?* (‚Was ist das?‘) sehr früh, ohne dessen innere Struktur analysiert zu haben – sie lernen ihn holistisch, da Erwachsene ihn mit hoher Frequenz verwenden (vgl.

²⁵ ‚Mit diesen Frequenzeffekten werden überzeugende Beweise für gebrauchsbasierte Spracherwerbsmodelle geliefert, welche die Rolle des Inputs betonen‘ (eigene Übersetzung).

²⁶ Es können auch kleinere sprachliche Einheiten wie Phone oder Silben gezählt werden, diese sind für die vorliegende Arbeit jedoch nicht relevant.

Lieven/Tomasello 2008: 174). Zudem tendieren hoch frequente Einheiten dazu, Irregularitäten zu bewahren, z.B. entstanden für weniger frequente unregelmäßige Verben im Englischen reguläre Alternativen (*weep – wept, weeped* ‚weinen‘), wohingegen hochfrequente Formen nicht angepasst wurden (*keep – kept, sleep – slept* ‚behalten‘; ‚schlafen‘). Zusätzlich werden sie oftmals phonetisch reduziert (vgl. Bybee 2008: 219).

Die Type-Frequenz gilt in der Regel als Indikator für die Produktivität einer Konstruktion, da unterschiedliche lexikalische Elemente in derselben Position verhindern, dass eine Konstruktion mit einem bestimmten lexikalischen Element assoziiert wird; da ein Mehr an Elementen zu allgemeineren Eigenschaften der Konstruktion führt, sodass diese auch eher neue Elemente adoptiert; und da mehr einsetzbare Elemente dazu führen, dass die Konstruktion öfter gebraucht wird und stärker im Gedächtnis verankert ist, sodass schneller auf sie zugegriffen werden kann (auch mit neuen Elementen) (vgl. Bybee/Thompson 1997: 384). So ermöglicht Type-Frequenz die (unbewusste) Analyse der Konstruktion. Beispielsweise wird das Perfekt im Deutschen mit den Hilfsverben *haben* oder *sein* gebildet, wobei weitaus mehr Verben das Hilfsverb *haben* verlangen. Mit anderen Worten, die Perfekt Konstruktion mit *haben* hat eine höhere Type-Frequenz als die Perfekt Konstruktion mit *sein*. Wittek und Tomasello (2005) zeigten, dass deutsch-sprechende Kinder das Perfekt eher variierten, d.h. mit neuen Verben bildeten, wenn die Konstruktion das Hilfsverb *haben* und nicht *sein* beinhaltete, also eine höhere Type-Frequenz aufwies. Bybee (2008: 221) hält fest: „a certain degree of type frequency is needed to uncover the structure of words and phrases“²⁷.

Auf Grund von Token- und Type-Frequenz entsteht also ein Nebeneinander von mehr oder minder lexikalisch spezifischen und mehr oder minder abstrakten Konstruktionen²⁸:

[...] as abstractions develop, they do not replace more item-based instances but [...] all remain in the inventory of constructions that speakers have created, to be drawn upon at various levels of abstraction depending on the context in which the utterance has to be under-

²⁷ ‚Type-Frequenz muss in einem gewissen Maße auftreten, um die Struktur von Wörtern und Phrasen aufzudecken‘ (eigene Übersetzung).

²⁸ Moders (1992, zitiert nach Bybee 2008: 222) Studie deutet darauf hin, dass kognitiv gefestigte, also hoch frequente Exemplare einer Konstruktion mitunter nicht mehr mit der abstrakten Konstruktion in Verbindung gebracht werden. Sie wählte unregelmäßige Verben des Englischen mit Vokalwechsel im Simple Past wie *string/strung* oder *write/wrote*. Dann erfand sie neue Verben und fragte Teilnehmende nach den Vergangenheitsformen. Sahen die Teilnehmenden zuerst ein mittelfrequentes Verb, waren die Teilnehmenden eher bereit, das Muster auf die neuen Verben zu übertragen als bei sehr frequenten Verben. Das bestätigt die Annahme, dass lexikalisch gefüllte Konstruktionen nicht von abstrakten Konstruktionen abgelöst werden, sondern nebeneinander existieren. Die Tatsache, dass zu hochfrequenten unregelmäßigen Verben im Englischen keine regulären Alternativen gebildet wurden, kann ebenfalls als Evidenz dahingehend gedeutet werden (vgl. Bybee 2008: 219).

stood or produced and the resulting ease of access to more or less item-based constructions. Thus, the inventory of constructions consists of constructions ranging from the fully lexically-specific through to the fully abstract[.]²⁹

(Lieven/Tomasello 2008: 175)

Erwerbsfehler von Kindern können so z.B. auf das Wechselspiel zwischen fest verankerten Chunks und aufstrebenden Regularitäten zurückgeführt werden. Ambridge und Rowland (2009) untersuchten positive Was-, Warum- und ja/nein-Fragen (*What does she want?* ‚Was möchte sie?‘) und deren Verneinungen (*What doesn't she want?* ‚Was möchte sie nicht?‘) von 50 Kindern im Alter von drei und vier Jahren und fanden heraus, dass Kinder, die positive Aussagesätze wie „she doesn't want“ korrekt produzieren, eher dazu tendieren, das Hilfsverb doppelt und somit inkorrekt zu gebrauchen, also z.B. *What does she doesn't want*. Das bedeutet, Kinder kombinieren das Schema „What does“ mit dem holistischen Chunk „She doesn't want“ (vgl. Bannard/Lieven 2012: 11).

3.1.2.2 KATEGORISIERUNG

Ferner wirkt sich die Nutzungshäufigkeit einzelner sprachlicher Elemente auf die Bildung von Kategorien aus, eine der grundlegendsten kognitiven menschlichen Fähigkeiten (vgl. Croft/Cruse 2004: 74). Abstrakte Kategorien entstehen durch den Vergleich konkreter Exemplare. Gemeinsamkeiten werden extrahiert, sodass Elemente, die sich auf eine bestimmte Art und Weise ähneln, zusammengefasst werden und eine abgestufte Struktur entsteht (Gentner/Medina 1998; vgl. Goldberg/Casenhiser 2008: 206). Je öfter ein Mitglied der Kategorie vorkommt, desto mehr trägt es zur Charakterisierung der Kategorie bei und desto eher wird es als Prototyp der Kategorie angesehen; Kategorien werden also nach Frequenz und nach Ähnlichkeit sortiert (Nosofsky 1988, vgl. Ellis/Ferreira-Junior 2009: 190). Kriterien können dabei phonetischer, semantischer, pragmatischer oder kontextueller Art sein (vgl. Bybee 2013: 54; vgl. Langacker 2013: 17).

Elio und Anderson (1984: 25) konnten zeigen, dass Testpersonen eine Kategorie besser erlernen, wenn sie anfangs nur einen Ausschnitt der Kategorie, welcher die frequentesten Mitglieder enthält, sehen: „an initial, low-variance sample of the most frequently occurring members may allow the learner to get a ‚fix‘ on what will account for most of the category

²⁹ ‚Abstraktionen, die nach und nach aus dem Input abgeleitet werden, ersetzen keine lexikalisch spezifischen Chunks. Alle zusammen bilden das Inventar an Konstruktionen eines Sprechers und können von diesem, je nach Kontext der zu verarbeitenden oder produzierenden Äußerung, auf unterschiedlichen Abstraktionsebenen abgerufen werden. Folglich reicht das Inventar an Konstruktionen von lexikalisch spezifisch bis zu gänzlich abstrakt.‘ (eigene Übersetzung).

members“³⁰. Analog fanden Goldberg und Casenhiser (2008; Goldberg et al. 2004) bezüglich des Spracherwerbs heraus, dass eine hohe Token-Frequenz eines einzelnen, allgemeinen Exemplars das Erlernen der abstrakten Konstruktion sowohl bei Kindern als auch bei Erwachsenen begünstigt. In einem Experiment präsentierten die Autoren den Teilnehmenden eine neue Konstruktion mit fünf unbekannten Verben. Anhand von Videoclips sollten die Teilnehmenden die Form der Konstruktion mit der entsprechenden Funktion verbinden. Bei einer Gruppe waren die unbekannten Verben ausgeglichen, mit relativ niedriger Token-Frequenz verteilt (1-1-2-2-2). In der anderen Gruppe waren dieselben Verben ungleich verteilt, d.h. ein Verb verfügte über eine weitaus höhere Token-Frequenz (4-1-1-1-1). Letztere Gruppe mit manipulierter Token-Frequenz erzielte bessere Ergebnisse, wobei auch die Gruppe mit gleichmäßiger Frequenz die Kontrollgruppe übertraf, d.h. „skewed frequency facilitates, but is not necessary for learning a new construction“³¹ (Goldberg/Casenhiser 2008: 207; Hervorh. im Original).

Matthews und Bannard (2010) untersuchten ebenfalls, wann Kinder Regularitäten aus lexikalischen Mustern ableiten und kamen zu etwas abweichenden Ergebnissen. Kinder sollten ihnen unbekannte Sequenzen, die von bereits bekannten Sequenzen in nur einem Wort abwichen (z.B. *A piece of toast* vs. *A piece of X*, z. B. *A piece of brick*) reproduzieren. Folgende Ergebnisse ließen sich festhalten: Kindern fiel es leichter, die neuen Sequenzen zu wiederholen, wenn (i) sich das vierte Wort nur schwer mit Hilfe der drei ersten Wörter vorhersagen ließ³² und (ii) das neue Wort vorherigen Varianten semantisch ähnelte, d.h. *slot entropy* (s. 5.2.1) und semantische Dichte beeinflussten den Erwerb der Konstruktion *A piece of X* ‚ein Stück X‘. Während bei Goldberg und Casenhiser (2008; Goldberg et al. 2004) eine niedrige *slot entropy*, also eine unausgeglichene Verteilung der Frequenzen, das Erlernen der Konstruktion begünstigte, sprechen Matthew und Bannards (2010) Ergebnisse dafür, dass eine hohe *slot entropy* den Erwerb der Konstruktion fördert.

Für erstere Annahme spricht, dass der tatsächliche Input, den Kinder erfahren, Goldberg und Casenhisers (vgl. 2008: 207) Design gleicht, d.h. eine Konstruktion tritt typischer-

³⁰ ‚ein anfängliches, kaum variierendes Beispiel der am häufigsten vorkommenden Mitglieder einer Kategorie befähigt Lernende, Charakteristika zu verinnerlichen, die für die meisten Mitglieder der Kategorie gelten‘ (eigene Übersetzung).

³¹ ‚manipulierte Frequenz vereinfacht das Erlernen einer neuen Konstruktion, ist aber nicht zwingend erforderlich‘ (eigene Übersetzung).

³² Dies wird mit dem statistischen Maß *entropy* erfasst, was angibt, wie gleichmäßig die möglichen Varianten verteilt sind: je ausgeglichener die Verteilung, d.h. je höher die *slot entropy*, desto weniger voraussagbar ist das vierte Wort (vgl. Gries/Ellis 2015: 8).

weise mit einem bestimmten Wort auf, auch wenn potentiell unterschiedliche Wörter die Leerstelle füllen können. Goldberg et al. (vgl. 2004: 297) zeigten bspw., dass im Bates Korpus (Bates et al. 1988), das an-das-Kind-gerichtete Sprache enthält, die Konstruktion *intransitive motion* ‚Bewegung, intransitiv‘ in 39 Prozent der Fälle mit dem Verb *go* ‚gehen‘ vorkommt, obwohl insgesamt 39 Verben in der Konstruktion verwendet werden. Dieses Erkenntnis deckt sich mit Zipfs Gesetz (1935), der bemerkte, dass viele Wörter selten und nur wenige Wörter häufig vorkommen, die dann eben einen Großteil der Tokens ausmachen. Dabei ist es logisch, Wörter wie *go* häufig zu verwenden, da *go* nicht so spezifisch wie z.B. *amble* ‚schlendern‘ ist und somit in vielen Kontexten eingesetzt werden kann (vgl. Goldberg et al. 2004: 298).

3.1.2.3 ZUVERLÄSSIGKEIT UND FORM

Neben der Frequenz spielt die Zuverlässigkeit der Form-Funktion-Beziehung beim Erlernen von Konstruktionen eine Rolle. Es geht darum, eine Assoziation zwischen einer Form bzw. einem Hinweis und der entsprechenden Funktion zu lernen, z.B. das Morphem „-s“ als Hinweis für eine Pluralform. Die Zuverlässigkeit der Form als Hinweis für eine bestimmte Interpretation bestimmt den Erwerb der Assoziation, mit anderen Worten, je zuverlässiger die Form eine Interpretation voraussagt, desto eher wird die Assoziation gelernt (vgl. Ellis 2008: 376). In aller Regel gibt es mehrere Hinweise, die auf dieselbe Bedeutung verweisen. Wie sich zeigte, konzentrieren sich Lerner gleichwohl zunächst nur auf einen Hinweis – dieser verfügt allerdings über die größte Validität, d.h. er kommt relativ oft vor und/oder ist relativ zuverlässig – nach und nach werden dann mehr Hinweise hinzugefügt (vgl. Ellis 2008: 378):

Cue validity is taken to be a function of both reliability and availability. Reliability refers to the extent to which you can depend on a cue to not mislead you. Availability refers to the extent to which a cue is there when you need it.³³

(MacWhinney et al. 1985: 205)

Neben der Validität scheint auch die Salienz oder *cue detectability* (vgl. ebd.; vgl. Ellis 2008: 378f.), also wie auffällig ein bestimmter Hinweis ist bzw. wie leicht er wahrgenommen werden kann, eine Rolle zu spielen. Elemente, die über eine höhere Anzahl an Phonemen verfügen, Elemente, die einen Vokal beinhalten (Syllabizität) und Elemente, die eine größere Sonorität aufweisen, sind dabei salienter als andere (vgl. Ellis 2007: 87). Wie bereits

³³ ‚Die Validität eines Hinweises setzt sich aus Reliabilität und Verfügbarkeit zusammen. Reliabilität bezeichnet, wie verlässlich ein Hinweis ist. Verfügbarkeit bedeutet, inwieweit ein Hinweis vorhanden ist, wenn er gebraucht wird.‘ (eigene Übersetzung).

erwähnt, werden Formen mit hoher Token-Frequenz oft phonetisch reduziert, sodass diese lediglich eine geringe physikalische Salienz aufweisen und so vor allem für L2-Lerner problematisch sind (vgl. Bybee 2008: 219; vgl. Ellis 2008: 380).

Lieven und Tomasello (2008: 181f.) nennen Komplexität als einen weiteren Faktor, der dazu beiträgt, wie leicht eine Konstruktion zu erwerben ist. Eine nicht analysierte Konstruktion ist dabei einfacher zu lernen als eine schematische Konstruktion mit einer oder mehreren Leerstellen. Darüber hinaus spielt es eine Rolle, ob neue Formen bereits bekannten Konstruktionen ähneln. Bspw. argumentieren Abbot-Smith und Behrens (2006), dass das Zustandspassiv im Deutschen vor dem Vorgangspassiv gelernt wird, da Kinder bereits einige Konstruktionen mit *sein* oder dem Partizip II kennen, nicht jedoch mit *werden*. Neben grammatischer Komplexität, die von der Anzahl der beteiligten Elemente einer Struktur abhängt, beeinflussen phonetische Komplexität – je bekannter ein Laut, desto leichter seine Produktion – und semantische Komplexität – je mehr Bedeutungen einer Struktur zugeordnet werden können, desto komplexer ist diese – den Erwerb eines sprachlichen Ausdrucks (vgl. Ellis 2007: 87; vgl. Fandrych/Tschirner 2007: 199).

Zusammenfassend lässt sich Goldschneider und DeKeyser's Studie (2001) zur Morphemerwerbsreihenfolge des Englischen anführen. Die Autoren konnten zeigen, dass die Inputfaktoren Frequenz, perzeptuelle Salienz und Zuverlässigkeit der Form-Funktions-Beziehung jeweils signifikant mit der Erwerbsreihenfolge korrelierten, d.h. die Faktoren konnten zusammengenommen einen Großteil der Erwerbsschwierigkeiten erklären (vgl. Ellis 2008: 381f.; vgl. Fandrych/Tschirner 2007: 199).

3.1.2.4 CROSS-LINGUISTISCHE VARIATION

Wenn Lernende Sequenzen also wiederholt im Input erleben müssen, um eine Verbindung zwischen Form und Bedeutung herstellen zu können, stellt sich die Frage, ob sich der Erwerb von Sprache zu Sprache unterscheidet. Ein Großteil der zitierten Studien bezieht sich auf den Erstspracherwerb des Englischen, was, im Gegensatz zum Deutschen, eine relativ strikte Satzstellung sowie weniger Variation in der Morphologie aufweist. Stoll et al. (2009) untersuchten an-das-Kind-gerichtete Sprache von englisch-, deutsch- und russischsprachigen Müttern und analysierten inwieweit diese einzelne Wörter und Wortsequenzen (zwei und drei Wörter) am Anfang einer Äußerung wiederholten. Dabei fanden sie heraus, dass die Sprache russischer und deutscher Mütter ähnlich zu der englischsprachiger Mütter ein hohes Maß an Wiederholung in satzanfänglicher Position aufwies. Englisch wies dabei

den höchsten Grad an struktureller Festigkeit auf, gefolgt von Deutsch und abschließend Russisch, was die strikte bis relativ freie Satzstellung der Sprachen sowie morphologische Unterschiede widerspiegelt. Selbst in Russisch betrug der Anteil an wiederholten Wörtern bzw. Worteinheiten 75 Prozent der Äußerungen, d.h. „even in some languages where word order is potentially free, the input children receive still contains a great degree of predictability at the beginnings of utterances“³⁴ (ebd.: 90). Außerdem wiederholten englischsprachige Mütter größere Einheiten, d.h. Zwei- oder Dreiworteinheiten, wohingegen im Russischen meist nur ein Wort anfänglich wiederholt wurde. Die Autoren stellten folglich die Frage, ob es so für russische Kinder schwerer sei, die Sprache zu erlernen, kamen jedoch zu dem Ergebnis, dass der Input genau die Informationen bereitstellt, die Kinder benötigen. Denn, während englische Kinder bspw. die Konstruktion *That's X* erwerben müssen, müssen russische Kinder lediglich die kürzere Einheit *Eto X* lernen (vgl. ebd.: 94).

Chan et al. (2009) untersuchten, wie erfolgreich englisch-, deutsch- und kantonesischlernende Kinder die semantischen Rollen Agens und Patiens in Transitivsätzen identifizierten. Die Autoren fokussierten dabei die Hinweise Wortstellung und Belebtheit und hielten fest, dass Wortstellung als Hinweis für die Agens-Patiens Beziehung im Englischen über eine höhere Validität als im Deutschen oder Kantonesischen verfügt, wobei Deutsch im Gegensatz zu Englisch zusätzlich den Kasus markiert. Die Eigenschaften [+belebt] für den Agens und [–belebt] für den Patiens gelten dagegen über Sprachgrenzen hinweg als relativ reliabel. Prototypische Sätze mit Belebtheit und Wortstellung als Hinweise wurden sehr leicht von den Kindern, unabhängig von ihrer Sprache, verstanden. Verwies allein die Wortstellung auf die Agens und Patiens Rollen, identifizierten englischlernende Kinder früher und öfter das erste Nomen als Agens, d.h. sie vertrauten am ehesten auf die Wortstellung als Hinweis für die Agens-Patiens Beziehung, was der hohen Validität im Input für Englisch entspricht.

Tomasello (vgl. 2003: 157) merkt für den Gebrauch der abstrakten Konstruktion Passiv an, dass Kinder, die bspw. Inuktitut, eine Nicht-Indoeuropäische Sprache, lernen, Passivsätze relativ früh im Gegensatz zu englischsprachigen Kindern produzieren. Auch dieses Erkenntnis lässt sich auf den Input zurückführen; zusätzlich spielt die Komplexität der Struktur eine Rolle:

³⁴ „In äusserungsinitialer Position ist der Input, den Kinder erfahren, selbst in Sprachen mit weitgehend freier Wortstellung überwiegend vorhersagbar.“ (eigene Übersetzung).

(1) Inuktitut passives are very common in child-directed speech, and (2) passive utterances are actually simpler than active voice constructions in Inuktitut because the passivized verb has to agree only with the subject, whereas the transitive verb has to agree with both subject and object[.]³⁵

(ebd.)

Festzuhalten ist, dass die Form-Bedeutungsbeziehungen, die Kinder erwerben, von konkret bis abstrakt, jeweils von den Eigenschaften der Sprache abhängen, die sie hören, was für einen gebrauchsbasierten Spracherwerb spricht.³⁶

3.2 CHUNKS IM ZWEITSPRACHERWERB

In diesem Kapitel soll der Frage nachgegangen werden, inwieweit sich Ergebnisse aus der L1-Erwerbsforschung auf den Zweitspracherwerb übertragen lassen. L2-Lernende erreichen normalerweise keine Kompetenzen, die mit muttersprachlichen Kompetenzen zu vergleichen wären, „they use too few native-like prefabs and too many foreign-sounding ones“³⁷ (Granger 1998: 158; vgl. Pawley/Syder 1983: 215). Nichtsdestoweniger bauen sie ebenfalls implizites Wissen über Laut- und Wortfolgen in der Fremdsprache auf. Jiang und Nekrasova (2007) testeten die Reaktionszeit von L2-Lernenden und von englischen Muttersprachlern, welche die Wohlgeformtheit von Sequenzen beurteilen sollten und fanden heraus, dass beide Gruppen schneller und mit weniger Fehlern auf formelhafte (z.B. *to tell the truth*) als auf nicht formelhafte Sequenzen (z.B. *to tell the price*) reagierten. Conklin und Schmitt (2008) zeigten, dass formelhafte Sequenzen sowohl von englischsprechenden Muttersprachlern als auch von L2-Lernenden schneller gelesen werden. Simpson-Vlach und Ellis (2009) und Ellis et al. (2008) untersuchten in einer Reihe psycholinguistischer Experimente, inwieweit korpusanalytische Maße wie Frequenz und *Mutual Information* (s. 5.2.2) implizit bei englischsprachigen Muttersprachlern sowie bei Nichtmuttersprachlern verankert sind und kamen zu dem Ergebnis, dass für L2-Lernende die Frequenz, für Muttersprachler jedoch der MI-Wert entscheidend für die Verarbeitung ist. Durrant und Schmitt (2009) stellten ebenfalls fest, dass Lerner dazu neigen, hochfrequente Kollokationen wie *good example* ‚gutes Beispiel‘ zu nutzen, während sie, im Vergleich zu Muttersprachlern, zu wenig Gebrauch von hochassoziativen Kollokationen wie z.B. *densely populated*

³⁵ ‚(1) Inuktitut Passivkonstruktionen kommen sehr häufig in an-das-Kind-gerichteter Sprache vor‘ und (2) Äußerungen im Passiv sind auf Inuktitut tatsächlich einfacher als Aktivkonstruktionen zu produzieren, da sich das passivierte Verb lediglich nach dem Subjekt, das transitive Verb dagegen nach Subjekt und Objekt richtet. (eigene Übersetzung).

³⁶ Hier wurden Studien fokussiert, die anhand von sprachspezifischen Konstruktionen gebrauchsbasierte Beschreibungsansätze legitimieren. Für weitere kontrastive Studien, vgl. Boas 2010.

³⁷ ‚sie nutzen zu wenig muttersprachenähnliche und zu viele unnatürlich klingende Wendungen‘ (eigene Übersetzung).

‚dicht bevölkert‘ machen.

Eine Studie von Gries und Wulffs (2005) deutet ebenfalls darauf hin, dass auch Fremdsprachenlerner Konstruktionen erwerben. Fortgeschrittene Englischlernende mit deutscher Muttersprache erhielten 16 Sätze, die vier verschiedene Verben (*cut, get, take, throw*) und vier verschiedene Argumentstrukturen (Ditransitiv-, Resultativ-, Transitiv- und *Caused-Motion*-Konstruktionen) enthielten. Die Teilnehmenden sollten die Karten nach ihrer allgemeinen Bedeutung sortieren und dafür vier verschiedene Stapel bilden. Folgendes Ergebnis ließ sich festhalten:

[...] foreign language learner subjects perceive sentences as much more similar to each other when they instantiate identical constructions than when they feature the same verbs [...] Apparently, then, the subjects relied more on the complement configurations and the corresponding semantics than on the superficial lexical similarity for sorting[.]³⁸
(ebd.: 194)

Das Ergebnis spricht also für die Annahme, dass auch L2-Lernende Regularitäten in Form von Konstruktionen aus dem Input ableiten.

Trotz eingeschränktem Gebrauch formelhafter Sprache konträr zu Muttersprachlern können Lernende frequente Sequenzen außerdem produktiv in ihrem Lernprozess nutzen. So liefern Langzeitstudien (Myles et al. 1999; Myles 2004) Hinweise darauf, dass Lernende durch formelhafte Chunks komplexe Strukturen, die über ihre Lernergrammatik hinausgehen, verwenden, wobei zwischen der holistischen Abrufung von Chunks, deren Analyse und der kreativen Sprachverwendung Wechselbeziehungen bestehen, die so einen positiven Effekt auf den Lernprozess erzielen:

It is as if, at any one time, learners are attempting to resolve the tension between complex but communicatively rich chunks on the one hand and simple but communicatively inadequate structures on the other hand. This is a dynamic tension that drives forward the overall development of the L2 system.³⁹

(Myles et al. 1999: 77)

³⁸ ‚Fremdsprachenlernende nehmen Sätze als ähnlicher in ihrer Bedeutung wahr, wenn diese dieselben abstrakten Konstruktionen und nicht dieselben Verben beinhalten. Für das Sortieren der Sätze vertrauten die Teilnehmenden also eher auf die Struktur und die entsprechende Semantik als auf oberflächliche, lexikalische Gemeinsamkeiten‘ (eigene Übersetzung).

³⁹ ‚Es ist, als ob Lernende stets versuchten, die Spannung zwischen komplexen aber kommunikativ wertvollen Chunks einerseits und einfachen aber kommunikativ unangemessenen Strukturen andererseits aufzulösen. Diese dynamische Spannung begünstigt die allgemeine Entwicklung der Lernergrammatik.‘ (eigene Übersetzung).

3.2.1 FAKTOREN IM GEBRAUCHSBASIERTEN ZWEITSPRACHERWERB

Während Frequenz, Zuverlässigkeit und Salienz sowohl für den Erst- als auch auf den Zweitspracherwerb relevant sind, lassen sich für den L2-Lernprozess einige zusätzliche Faktoren – limitierter Input, selektive Wahrnehmung und Interferenz – identifizieren, die das Fremdsprachenlernen erschweren. Das heißt, auch wenn L2-Lerner ebenfalls probabilistische Informationen über die Fremdsprache speichern und Konstruktionen lernen, können diese Faktoren erklären, warum der gebrauchsbasierte Zweitspracherwerb in der Regel nicht zu nahezu muttersprachlichen Kenntnissen führt (vgl. Ellis 2008: 382f.).

3.2.1.1 LIMITIERTER INPUT

In gebrauchsbasierten Modellen lässt sich der Spracherwerb im Grunde als „sampling problem“ (Ellis et al. 2015: 364) beschreiben, d.h., den Lernenden muss eine genügend große Stichprobe an relevantem, authentischem Input zur Verfügung stehen, sodass sie auch weniger frequente Ausdrücke wahrnehmen:

Becoming fluent requires a sufficient sample of needs-relevant authentic input for the necessary implicit tunings to take place. The ‚two puzzles for linguistic theory‘, nativelike selection and nativelike fluency (Pawley & Syder, 1983), are less perplexing when considered in these terms of frequency and probability. There’s a lot of tallying to be done here. The necessary sample is certainly to be counted in terms of thousands of hours on task.⁴⁰
(Ellis 2009: 152)

Kinder verfügen über diese tausende Stunden an Input und an Interaktion, wohingegen erwachsene Lernende nur einen sehr limitierten Zugang zu authentischer Sprache haben. Dies gilt vor allem für gesteuerte Erwerbssituationen. Input bezieht sich hierbei auf bedeutungsvolle sprachliche Äußerungen in der Zielsprache, denen ein Lerner in einem ebenfalls bedeutungsvollen Kontext begegnet (vgl. Barcroft/Wong 2013: 627). Der Input muss bedeutungsvoll und verständlich sein, da Lernende nur so Form-Bedeutungsverbindungen herstellen und in letzter Konsequenz die Sprache erwerben können (vgl. ebd.: 635). Alle im Input enthaltenen Informationen können von den Lernenden indes nicht verarbeitet werden, sodass zwischen *Input* und *Intake* – Informationen, welche die Lernenden tatsächlich nutzen – unterschieden wird (Corder 1967). Faktoren, die einen Einfluss darauf haben, welche Informationen Lernende fokussieren und welche sie vernachlässigen, werden nach-

⁴⁰ ‚Um eine Sprache fließend zu beherrschen, benötigen Lernende eine genügend große Stichprobe an relevantem, authentischem Input, damit verarbeitende Effekte [bspw. Chunking und die Bildung abstrakter Kategorien] erfolgen können. Die zwei sprachtheoretischen Rätsel *nativelike selection* und *nativelike fluency* erscheinen vor dem Hintergrund von Frequenz und Wahrscheinlichkeit weniger verblüffend. Eine sehr große Datenmenge muss implizit vom Lernenden ausgewertet werden – die Zeit dafür ist sicherlich in Tausenden von Stunden zu messen.‘ (eigene Übersetzung).

folgend expliziert.

3.2.1.2 TRANSFER

Die Tatsache, dass Lernende bereits ihre Muttersprache und eventuell andere Sprachen beherrschen, lässt sie unbewusst nach bestimmten sprachlichen Hinweisen suchen, während andere nicht wahrgenommen werden. Mit anderen Worten, bereits erworbene Bedeutungen und Assoziationen führen zu einer angelernten selektiven Aufmerksamkeit bzw. zu perzeptivem Transfer (vgl. Ellis 2008: 388f.; vgl. Fandrych/Tschirner 2007: 200)⁴¹. Bspw. wird das Tempus redundant sowohl durch korrekte Tempusformen der Verben als auch durch Adverbien ausgedrückt. Während L1-Lernende die Bedeutung von Zeitadverbien relativ spät erlernen, verwenden L2-Lernende anfangs ausschließlich Adverbien, um Zeitlichkeit auszudrücken. Da sie den Einsatz von temporalen Adverbien, die zudem eine hohe Salienz und Zuverlässigkeit aufweisen, bereits aus ihrer Muttersprache als erfolgreiche Strategie kennen, um Temporalität zu kennzeichnen, werden temporale Adverbien von L2-Lernenden beachtet und genutzt, wohingegen phonetisch reduzierte und damit unauffällige Tempusmarkierungen ignoriert werden (Bardovi-Harlig 1992; vgl. Ellis 2007: 83). Analog beschreibt Pica (1983), wie L2-Lernende, die die Sprache ungesteuert erwerben, das Plural „-s“ bei Nomen weglassen, wenn die Mehrzahl bereits durch ein (auffälliges) Indefinitum markiert wurde, also z.B. *a few book (vgl. Ellis 2008: 389).

Ferner scheint der Erwerb der Lese- und Schreibfähigkeit zur selektiven Aufmerksamkeit beizutragen. Wray (vgl. 2002: 200) geht davon aus, dass L2-Lernende den Fokus beim Analysieren nicht auf enthaltene grammatische Informationen, sondern auf einzelne Wörter legen, weil sie als Jugendliche bzw. Erwachsene bereits Lesen und Schreiben können, was die Aufmerksamkeit auf einzelne Wörter fördert: „It seems to be only in the context of reading and writing that the whole vocabulary is assigned word boundaries“⁴² (ebd.: 137, vgl. auch Aguado 2013: 137f.). Yorio (1989) untersuchte eine Gruppe Englischlernende in den USA in Bezug auf die Verwendung formelhafter Sprache beim Schreiben. Sie stellte fest, dass die Lernenden versuchten, formelhafte Sequenzen zu verwenden, diese aber fehlerhaft produzierten, z.B. *take advantages of, being taking care of, make a great influence*

⁴¹ Natürlich übernehmen L2-Lernende auch produktiv Strukturen aus ihrer Muttersprache oder aus anderen bereits erlernten Sprachen und übertragen diese in die Zielsprache, was zu positivem oder negativem Transfer (Interferenz) auf allen linguistischen Ebenen – von der Phonetik bis zur soziolinguistisch-interkulturellen Ebene – führt (vgl. Odlin 2003: 436).

⁴² ‚Es scheint, als ob das Vokabular erst im Prozess der Vermittlung der Lese- und Schreibfähigkeit in Wörter unterteilt wird.‘ (eigene Übersetzung).

(ebd.: 62f.). Dass diese Wortverbindungen den korrekten Sequenzen sehr ähneln, deutet darauf hin, dass sie von den Lernenden als Einheiten wahrgenommen wurden. Zudem müssen sie analysiert worden sein, da sie nicht einfach imitiert, sondern verändert wurden. Versuchen Lernende, die Sequenz zu einem späteren Zeitpunkt zu reproduzieren, setzten sie die einzelnen Lexeme gemäß ihrer Lernaltersprachengrammatik zusammen, was zu formseitig fehlerhaften Äußerungen führen kann.

VanPatten (2004: 7) macht ebenfalls Probleme bei der Inputverarbeitung bzw. die Tatsache, dass Lernende den Input auf Grund der limitierten Kapazität des Arbeitsgedächtnisses nur beschränkt nutzen können (Intake) für Erwerbsschwierigkeiten verantwortlich und formuliert allgemeine Prinzipien, welche die Verarbeitung steuern. Das erste Prinzip und die sich daraus ableitenden Subprinzipien können unter anderem die eben angeführten Beispiele erklären.

Principle 1. The Primacy of Meaning Principle. Learners process input for meaning before they process it for form.⁴³

(ebd.: 14; Hervorh. im Original)

Folgende Unterprinzipien ergeben sich aus dem Prinzip „Bedeutung vor Form“: Inhaltswörter werden zuerst, lexikalische Einheiten vor grammatischen, nicht-redundante bedeutungstragende grammatische Einheiten vor redundanten bedeutungstragenden Einheiten und bedeutungstragende grammatische Einheiten vor nicht bedeutungstragenden Einheiten (unabhängig von Redundanz) verarbeitet (vgl. ebd.). Die Annahme, dass Lernende primär Inhaltswörter und lexikalische Einheiten fokussieren, liefert also eine Erklärung für die Tatsache, dass sie bspw. Tempus- und Pluralmarkierungen übergehen. Ein weiteres Prinzip beschreibt die Tendenz von Lernenden, das erste vorkommende Nomen oder Pronomen als Subjekt bzw. als Agens zu identifizieren (vgl. ebd.: 15)⁴⁴.

Das nächste Kapitel befasst sich mit den Konsequenzen, die sich aus diesen Prinzipien ergeben, d.h. mit der Frage, wie der Input für eine optimale Verarbeitung aufbereitet werden muss.

⁴³ ‚Prinzip 1. Das *Primacy of Meaning Principle*. Lernende verarbeiten den Input inhaltlich bevor sie ihn formseitig analysieren.‘ (eigene Übersetzung).

⁴⁴ Empirische Belege für VanPattens Model sind in Barcroft und Wong (2013: 639-641) aufgeführt.

3.2.2 IMPLIKATIONEN FÜR DEN GESTEUERTEN ZWEITSPRACHERWERB

Fremdsprachenlernende verfügen also wie L1-Sprecher über statistisches Wissen in Bezug auf sprachliche Einheiten in der L2 und beziehen formelhafte Sprache produktiv in den Erwerbsprozess ein (vgl. Ellis et al. 2015: 375). Die einheitenbildenden Sequenzen befähigen sie unter anderem, komplexe Strukturen zu produzieren, die den Stand ihrer Lernergrammatik übersteigen. Dabei nutzen sie insgesamt jedoch eine kleinere Anzahl formelhafter Sequenzen als Muttersprachler. Vor allem weniger häufige aber hochassoziative Wortketten werden von L2-Lernenden oft nicht als Einheiten wahrgenommen. Neben dem, im Vergleich zu L1-Lernenden, stark begrenzten Input, erschwert besonders die Erstsprache der Lernenden den Erwerb von Konstruktionen, was sich in Interferenzerscheinungen und perzeptivem Transfer äußert.

Allgemein erscheint es sinnvoll und effektiv, frequente Wörter oder Wortverbindungen vor weniger häufigen Wörtern zu lernen, da sie den meisten Nutzen bringen (vgl. Tschirner 2005: 135). Neben der Token-Frequenz spielt auch die Type-Frequenz von Mehrwortmustern eine Rolle, denn „if a learner is to apply a pattern productively to forms not necessarily encountered before, he/she must have encountered the pattern with a number of different items in it“⁴⁵ (Bybee 2008: 221). Gemäß Goldberg und Casenhiser (vgl. 2008: 210) Ergebnissen erscheint es förderlich für den Erwerb einer Konstruktion zu sein, den Input anfänglich übermäßig mit einem hochfrequenten Exemplar, das prototypisch für die Bedeutung der Konstruktion, gleichzeitig aber allgemein und somit vielseitig anwendbar ist, zu durchsetzen und erst im Anschluss andere seltenere Typen zu vermitteln. Ellis und Ferreira-Junior (2009) konnten für den ungesteuerten Zweitspracherwerb bestätigen, dass der Erwerb sprachlicher Konstruktionen von ihrer Frequenz, der Type/Token-Verteilung und der Prototypizität sowie der Vagheit ihrer Bedeutung abhängt. Für die Anwendung im Fremdsprachenunterricht stellt sich allerdings die Frage, inwieweit es möglich ist, das Sprachangebot immer wieder mit demselben Chunk zu durchsetzen, denn „focused training on a narrow subtype of a pattern could lead to excessive boredom“⁴⁶ (Goldberg/Casenhiser 2008: 210).

⁴⁵ ‚Soll ein Lerner ein Muster produktiv auf Formen übertragen, denen er möglicherweise im Input noch nicht begegnet ist, muss er das Muster mit verschiedenen Füllern gelesen/gehört haben.‘ (eigene Übersetzung).

⁴⁶ ‚Fokussiertes Einüben eines einzelnen Untertyps eines Musters könnte unglaublich langweilig werden.‘ (eigene Übersetzung).

Weitere Maßnahmen für eine lernförderliche Aufbereitung des Inputs werden unter dem Begriff *input enhancement* (Sharwood Smith 1993) subsumiert. Gemeint ist, dass, nachdem der Input inhaltlich erfasst wurde, sprachliche Details hervorgehoben werden, um die Aufmerksamkeit der Lernenden auf diese zu lenken (vgl. Tschirner 2000: 74). Bspw. können wenig auffällige grammatische Informationen **vergrößert**, unterstrichen, **fett** oder *kursiv* gedruckt oder **farbig** markiert werden. Neben diesen Techniken kann der Input massiv mit der gewünschten zielsprachigen Form durchsetzt werden, um von den Lernenden wahrgenommen zu werden. Dieses Verfahren wird in der Forschung als Input-Flut bezeichnet.

Neben frequentem und strukturiertem Input sowie expliziter Grammatikvermittlung als Voraussetzungen für einen produktiven Umgang mit Chunks (vgl. Aguado 2013: 138) wirft Rostila (vgl. 2012: 224) die Frage auf, wie lange Lernende dem Input ausgesetzt werden sollten, bis sie selbst eine bestimmte Konstruktion produzieren sollen.

4. ANSÄTZE ZUR VERMITTLUNG VON CHUNKS

Wie in den vorherigen Kapiteln gezeigt wurde, spielt Chunking eine wesentliche Rolle in der Sprachverwendung sowie beim Spracherwerb. Wird dies berücksichtigt, scheinen Chunks besser als Wörter dazu geeignet, als Basiseinheiten in der Fremdsprachenvermittlung zu dienen. Darüber hinaus wird die traditionelle Trennung von Lexikon und Grammatik in Frage gestellt. Im Folgenden sollen zwei Ansätze präsentiert werden, die zeigen, wie diese Erkenntnisse in der Praxis umgesetzt werden können. Dabei fokussiert der Lexical Approach den kommunikativen Mehrwert, der sich aus dem Chunk-Lernen ergibt, d.h. Chunks befähigen Lernende, flüssiger und natürlicher in der Fremdsprache zu kommunizieren. Der *Chunk-und-Konstruktionen-Ansatz* (Handwerker 2008), auf der anderen Seite, hat zum Ziel, Lernende beim Aufbau einer impliziten Grammatik durch die Extraktion von frequenten, abstrakten Konstruktionen im Input zu unterstützen.

4.1 THE LEXICAL APPROACH

Nattinger und DeCarrico (1992) bezeichnen Chunks als *lexical phrases*, sprachliche Muster, die sie aus Korpora gesprochener und geschriebener Sprache extrahieren. *Lexical phrases* werden dabei als „multi-word lexical phenomena [...], conventionalized

form/function composites that occur more frequently and have more idiomatically meaning than language that is put together each time“⁴⁷ definiert (ebd.: 1). Darüber hinaus betonen die Autoren, dass *lexical phrases* pragmatische Funktionen ausüben, d.h. sie beleuchten neben formalen auch funktionale Aspekte. Dabei teilen sie die *lexical phrases* in drei Gruppen ein, die sie als *social interactions*, *necessary topics* und *discourse devices* bezeichnen (vgl. ebd.: 59-65). Unter *social interactions* werden Phrasen gefasst, die soziale Beziehungen markieren, bspw. *I didn't catch your name* oder *(please,) if you don't mind*. *Necessary topics* sind Themen, über die Lernende oft sprechen müssen, bzw. frequente Alltagsthemen. Beispiele sind *I don't speak _____ very well; is it going to [rain, snow] X?; it (doesn't) fit(s)*. Zuletzt dienen *discourse devices* der Strukturierung von Diskursen, z.B. *as a result (of X); it seems (to me) that X*.

Lewis (2008: 3) fasst den Lexical Approach folgendermaßen zusammen: „language consists not of traditional grammar and vocabulary but often of multi-word prefabricated chunks“⁴⁸. Dabei unterteilt er Chunks in vier Typen: *words*, *collocations*, *fixed expressions* und *semi-fixed expressions*. Diese müssen jedoch nicht wie bei Nattinger und DeCarrico pragmatische Funktionen ausführen. Grundsätzlich plädiert Lewis (2008: 9) dafür, größere Chunks zu unterrichten, als dies bisher der Fall ist, z.B. *suspicious of people who...* anstelle *suspicious of...* oder *relevant to our discussion/problem/needs* anstatt *relevant to...* Für die Umsetzung des Lexical Approach in der Praxis formuliert er einige Grundprinzipien (vgl. ebd.: 49-54). So soll der Fokus z.B. vermehrt auf dem Input liegen. Darüber hinaus sollen Chunks öfter wiederholt werden, da nicht davon ausgegangen werden kann, dass sie ab dem ersten „Antreffen“ beherrscht werden. Ein erster Schritt, Chunks zu lernen, ist diese wahrzunehmen und zu identifizieren. Schließlich sollen Lernende auf Übergeneralisierungen hingewiesen werden, die unvermeidbar scheinen. Diese Basisprinzipien spiegeln sich auch in den Aufgabentypen wider, die Lewis (2008: 89-91) präsentiert: *identifying chunks*, *matching*, *completing*, *categorising*, *sequencing* und *deleting*. Chunks sollen also z.B. zugeordnet werden (*matching*), wie bei *every now and.../once in a blue.../up to... und now/moon/then/* oder aussortiert werden (*deleting*), wie bei *BRIGHT – idea, green, ~~smell~~, child, day, room* (vgl. ebd.: 92, 94).

⁴⁷ ‚Polylexikale Phänomene, konventionalisierte Form-Funktionspaare, die häufiger vorkommen und eine höhere Idiomatizität aufweisen, als sprachliche Elemente, die neu zusammengesetzt werden‘ (eigene Übersetzung).

⁴⁸ ‚Sprache besteht nicht aus traditioneller Grammatik auf der einen und Wortschatz auf der anderen Seite, sondern meist aus vorgefertigten polylexikalischen Chunks‘ (eigene Übersetzung).

Dem Lexical Approach folgend sollte demnach z.B. die Form *capable of* oder auch *capable of something* anstelle von *capable* unterrichtet werden. Laut Holme (2010: 358) ist dies insofern problematisch, als es den Lernenden nicht genügend Informationen zur Verfügung stellt, „because it under-specifies the productivity of the form being used“⁴⁹. Anstelle der Sequenz *capable of something* sollte eine Konstruktion vermittelt werden, die lediglich das Wort *of* enthält und anhaltende, auf Erkenntnis gründende Zustände beschreibt, also eine Reihe an Füllern wie *aware*, *conscious*, *suspicious* und *certain* zulässt⁵⁰. Ein kommunikativer Ansatz wie der Lexical Approach nutzt demnach nicht das ganze Potenzial, welches Chunk-Lernen für den Einsatz im Fremdsprachenunterricht bereithält. Um der Tatsache, dass Lernende sich auch auf höheren Niveaus auf lexikalische Chunks stützen, entgegenzuwirken, sollten schematische Konstruktionen fokussiert werden: „[This] is exactly why constructions need to be explored more from the perspective of their schematicity and less from that of their existence as fixed expressions“⁵¹ (ebd.: 363). Holme (2010) untersuchte in einem Experiment, wie sich das explizite Unterrichten von produktiven Konstruktionen auf den Lernerfolg auswirkt und kam zu dem Ergebnis, dass die entsprechenden Konstruktionen schneller erlernt wurden und die Teilnehmenden schriftlich weniger Fehler machten. Aufgabentypen beinhalteten u.a. die Suche nach alternativen Formen für eine Konstruktion (z.B. *People who live in polluted places may be risking their health/Children who play computer games a lot could be risking their sight* ‚Menschen, die in verseuchten Gebieten leben, riskieren möglicherweise ihre Gesundheit/Kinder, die oft Computerspiele zocken, riskieren möglicherweise ihr Augenlicht‘ (eigene Übersetzung)) und die Suche nach prototypischen Realisationen (ebd.: 371). Schwierigkeiten beim Unterrichten von Konstruktionen sieht er zum einen in der Annahme, dass nicht-muttersprachliche Lehrende diese nur zögerlich verändern, sodass ein Inventar an Konstruktionen für bestimmte Bereiche und Themen erarbeitet werden muss. Darüber hinaus müssen Unterrichtsmethoden für alle Niveaus und Altersgruppen entwickelt werden, „mingling the conscious explorations

⁴⁹ ‚weil die Produktivität der verwendeten Form nicht deutlich genug zum Ausdruck kommt‘ (eigene Übersetzung).

⁵⁰ M.E. stellen die unterschiedlichen syntaktischen Verwendungsweisen ein Problem in der Vermittlung der Konstruktion dar, vgl. *capable of sth/doing sth*; *aware of sth/become aware of someone doing sth*; *suspicious of someone/sth*; *certain of sth*.

⁵¹ ‚Aus diesem Grund sollten Konstruktionen eher schematisch und weniger ausgehend von ihrem Status als feste Ausdrücke analysiert werden.‘ (eigene Übersetzung).

of meaning that are applicable to the analytic, educated learners [...] with the more intuitive usage-based approaches that will work better with lower age groups“⁵² (ebd.: 373).

Nachfolgend soll Handwerkers (2008) *Chunk-und-Konstruktionen-Ansatz* präsentiert werden, der sich ebenfalls mit abstrakten Konstruktionen und mit der Frage beschäftigt, wie im Input bereitgestellte Sequenzen nutzbar gemacht werden können, sodass Lernende abstraktes grammatisches Wissen entwickeln.

4.2 DER CHUNK-UND-KONSTRUKTIONEN-ANSATZ

Die grundlegenden Annahmen, die bis hierhin diskutiert wurden, d.h.,

dass (i) implizites grammatisches Wissen auf dem Weg des Speicherns von Chunks erworben wird [...], dass (ii) massenhafter Input mit situationseingebetteten Chunk-Angeboten gleicher Struktur die grammatische Abstraktion befördert [...], dass (iii) erst eine explizite Grammatikinstruktion das Aufbrechen der Chunks bei erwachsenen Fremdsprachenlernern garantiert und damit die gefrorene Grammatik in den Chunks für die Lernergrammatik verfügbar macht

(Handwerker 2015: 129)

bilden unter anderem die Basis für Handwerkers (2008) *Chunk-und-Konstruktionen-Ansatz*, der einen Versuch darstellt, die sprach- und lerntheoretischen Erkenntnisse für das eigenständige Lernen und den Unterricht nutzbar zu machen. Handwerker (2015: 129) bezieht sich in ihrem Ansatz auf das Beschreibungsmodell der Konstruktionsgrammatik, da es nicht, wie andere Modelle „mit Regeln über atomaren Einheiten operiert [und so] dem holistischen Verarbeiten von komplexen Einheiten entgegentläuft“. Abbildung 3 zeigt, wie sich Lerntheorie und Konstruktionsgrammatik ergänzen.

Lerntheorie	Grammatikmodell
C Chunk-Angebot für den Lerner:	K Konstruktionen als Form-Bedeutungspaare:
C1 Strukturierter Input mit salienten Ausdruckssequenzen gleicher Struktur und im Kontext herleitbarer Bedeutung	K1 Präsentation von Mustern und Konstruktionsbedeutungen
Einzelbeispiele <ul style="list-style-type: none"> • Er läuft seine Schuhe schief • Sie hungert sich krank • Sie schwebt über den Laufsteg • Er stolpert auf das Fußballfeld 	Exemplarisch <ul style="list-style-type: none"> • Typen von Resultativkonstruktionen im L2-Lernformat • Manner-of-Motion-Verb-Konstruktionen im L2-Lernformat
C2 <ul style="list-style-type: none"> • Tuning durch Input-Flut: Aufbau von Erwartungen an den Input • Training mit bedeutungsbasierter Formfokussierung 	K2 <ul style="list-style-type: none"> • Instruktion zur Identifikation von Konstruktionen • Instruktion zur gefrorenen grammatischen Information

⁵² ‚indem explizite Herangehensweisen zur Bedeutungerschließung, die sich für analytische, lernerfahrene Gruppen eignen, mit eher intuitiv ausgerichteten, gebrauchsbasierten Ansätzen, die für jüngere Lernende geeignet sind, gemischt werden‘ (eigene Übersetzung).

Lerntheorie	Grammatikmodell
C3 Lernalterssprachliche Produktion durch <ul style="list-style-type: none"> • die Kombination von Vorgefertigtem und die Variation lexikalischer Füllungen von Chunks • die parallele Nutzung von Wissen zu syntaktischer und semantischer Valenz lexikalischer Einheiten 	K3 Unterstützung der lernalterssprachlichen Produktion durch <ul style="list-style-type: none"> • die Instruktion zu Lexem-Konstruktion-Affinitäten und zu regelhaften und gebrauchsbedingten Beschränkungen • die Instruktion zur parallelen Anwendung syntaktischer Regeln

Abbildung 3. Komponenten des Chunk-und-Konstruktionen-Ansatzes.

(ebd.: 131)

In der Anwendung stellt sich die Frage, wie Konstruktionen für die L2-Instruktion präsentiert werden können. Wichtig scheint einerseits, Merkmale einer Konstruktion offenzulegen, die es Lernenden erlauben, die Konstruktion im jeweiligen situativen Kontext zu verstehen und andererseits, bezüglich der Produktion, Restriktionen der Elemente, mit denen eine Konstruktion gefüllt werden kann, aufzuzeigen (vgl. ebd.: 135f.). Auf Grund des limitierten authentischen Inputs, den L2-Lernende erfahren, plädiert Handwerker (vgl. ebd.: 137) für den Einsatz negativer Evidenz im gesteuerten Fremdsprachenerwerb, wobei sich (Nicht-) Akzeptabilitätsurteile mitunter von Muttersprachler zu Muttersprachler unterscheiden. Die Autorin (ebd.: 149) schlussfolgert,

[...] dass alles, was für den nativen Sprecher interpretierbar ist, auch dem fortgeschrittenen Lerner verfügbar gemacht werden sollte, denn einzelne strikte Urteile der Nicht-Akzeptabilität haben keinen Einfluss darauf, was den Lerner im fremdsprachlichen Input erwartet.

Um bevorzugte Verwendungsweisen und Einschränkungen im Gebrauch von Konstruktionen offenzulegen, können Methoden der Korpuslinguistik eingesetzt werden. Dies soll im folgenden Kapitel genauer erläutert werden.

5. KORPUSLINGUISTIK

Die Korpuslinguistik bezeichnet eine sprachwissenschaftliche Methode, bei der sprachliche Elemente aus einem Korpus extrahiert und ausgewertet werden. Ein Korpus bezeichnet dabei eine Sammlung von authentischen schriftlichen oder gesprochenen Texten, die in der Regel in maschinenlesbarer Form vorliegen und die nach Möglichkeit repräsentativ und ausgewogen in Bezug auf eine bestimmte sprachliche Varietät oder ein bestimmtes sprachliches Register oder Genre sind. *Repräsentativ* meint in dem Zusammenhang, dass alle Aspekte einer sprachlichen Varietät im Korpus wiederzufinden sind; *ausgewogen* meint, dass diese Aspekte zu gleichen Teilen, wie sie in der Varietät vorkommen, auch im Korpus enthalten, also proportional repräsentiert sind (vgl. Gries 2008b: 411). Damit soll unter

anderem der Kritik Rechnung getragen werden, Korpora seien stets unvollständig und könnten eine unendliche Sprache nicht erfassen (vgl. McEnery/Wilson 2001: 30). Während bspw. Chomsky davon ausgeht, dass introspektive Verfahren eine geeignetere Methode zur Analyse der Kompetenz darstellen, haben Korpusdaten auf der Performanzebene den Vorteil, dass sie beobachtbar und jederzeit überprüfbar sind (vgl. ebd.: 14). Außerdem zeigte sich: „human intuition about language is highly specific, and not at all a good guide to what actually happens when the same people actually use the language“⁵³ (Sinclair 1991: 4). Für die vorliegende Arbeit besonders relevant ist zudem die Tatsache, dass die Frequenz bei der Introspektion ignoriert wird, denn Menschen haben kaum ein Gespür dafür, wie oft ein bestimmtes sprachliches Element vorkommt: „Human beings have only the vaguest notion of frequency of a construct or word. Natural observation of data seems the only reliable source of evidence for such features as frequency“⁵⁴ (McEnery/Wilson 2001: 15).

Wenn wir also bedenken, dass die in den vorangegangenen Kapiteln vorgestellten Sprach(erwerbs)theorien die Rolle des tatsächlichen Inputs im Spracherwerb betonen, wobei den Frequenzeffekten – Token-Frequenz als Maß für die kognitive Verankerung und Type-Frequenz als Maß für die Produktivität einer Konstruktion – ein besonderer Status zukommt und wenn wir die Tatsache berücksichtigen, dass die Korpuslinguistik fast ausschließlich mit Frequenzanalysen arbeitet, erscheint es naheliegend, dass korpuslinguistische Methoden besonders im Rahmen der Kognitiven Linguistik Anwendung finden (vgl. Gries 2008b: 411). Zudem geht auch die Korpuslinguistik davon aus, dass Lexik und Grammatik ineinander greifen und sich gegenseitig beeinflussen (vgl. Fandrych/Tschirner 2007: 197). Folglich eignet sich die Korpuslinguistik in besonderem Maße als Methode, um in der vorliegenden Arbeit rekurrente Chunks in der Lernersprache zu extrahieren. Grundlage hierfür bieten Lernerkorpora, d.h. von Lernenden produzierte Sprachdaten.

⁵³ ‚Menschliche Intuition bezüglich Sprache ist äußerst individuell und stimmt kaum damit überein, was passiert, wenn dieselben Menschen die Sprache tatsächlich gebrauchen.‘ (eigene Übersetzung).

⁵⁴ ‚Menschen haben lediglich eine äußerst vage Vorstellung von der Frequenz eines Konstrukts oder eines Wortes. Die Analyse von Daten kann also als einzig zuverlässige Beweisquelle für Merkmale wie Frequenz gelten.‘ (eigene Übersetzung).

5.1 LERNERKORPORA

Im Bereich des Zweitspracherwerbs sowie in der Fremdsprachendidaktik können insbesondere Lernerkorpora neue Informationen und Erkenntnisse liefern. Ein Lernerkorpus bezeichnet dabei ebenfalls eine elektronische Sammlung authentischen Sprachgebrauchs, die Texte wurden jedoch von Lernenden, welche die jeweilige Sprache als Fremd- oder Zweitsprache erwerben, produziert. Die Daten eines Korpus' werden, je nach Ziel der Untersuchung, nach bestimmten Kriterien, z.B. Niveau der Teilnehmenden zusammengetragen (vgl. Granger 2002: 4). Gerade im gesteuerten Zweitspracherwerb stellt sich die Frage, inwieweit der Sprachgebrauch tatsächlich als *authentisch* bezeichnet werden kann. Wichtig erscheint, dass es sich um längere Texte und nicht etwa um einzelne Wörter oder Sätze handelt. Zudem hält Granger (2002: 5) fest:

In relation to learner corpora the term 'authentic' therefore covers different degrees of authenticity, ranging from 'gathered from the genuine communications of people going about their normal business' to 'resulting from authentic classroom activity'. In as far as essay writing is an authentic classroom activity, learner corpora of essay writing can be considered to be authentic written data, and similarly a text read aloud can be considered to be authentic spoken data.⁵⁵

Während es für das Englische eine ganze Reihe von Lernerkorpora gibt, existieren für das Deutsche nur wenige frei zugängliche Lernerkorpora. Die Website „Learner corpora around the world“ (<<https://www.uclouvain.be/en-cecl-lcworld.html>>, 23.11.16), die von der Universität Louvain betrieben wird, verzeichnet 152 Lernerkorpora, 96 haben als Zielsprache Englisch. Für das Deutsche sind als öffentlich zugängliche Korpora Das Corpus of Learner German (CLEG13), das deL1L2IM Korpus, das Falko Korpus (Fehlerannotiertes Lernerkorpus), das als Datengrundlage für die Analyse in 6. dient, das Lernerkorpus LeKo und das MERLIN Korpus (Multilingual Platform for the European Reference Levels: Exploring Interlanguage in Context) aufgeführt.

⁵⁵ „In Bezug auf Lernerkorpora können dem Begriff 'authentisch' unterschiedliche Grade an Authentizität zugeordnet werden, die von 'in natürlichen Konversationen von Menschen, die ihrer Routine nachgehen, gesammelt' bis zu 'in authentischen Unterrichtssituationen entstanden' reichen. Insofern, als das Schreiben eines Essays eine authentische Unterrichtsaktivität darstellt, können Lernerkorpora, die aus Essays bestehen, als authentische, schriftliche Daten betrachtet werden. Analog können laut vorgelesene Texte als authentische, gesprochene Daten gelten.“ (eigene Übersetzung).

5.2 KORPUSANALYTISCHE METHODEN ZUR IDENTIFIKATION VON CHUNKS

Kapitel 3 beschäftigte sich mit Lernmechanismen bzw. Inputfaktoren, die einen Einfluss auf den Erwerb von festen bis mehr oder minder abstrakten Konstruktionen haben. Um diese Faktoren zu untersuchen, müssen sie messbar gemacht werden (vgl. Gries/Ellis 2015: 2). Nachfolgend werden korpusanalytische Methoden zur Identifikation von Chunks sowie Methoden, die versuchen, psychologische Faktoren statistisch zu erfassen, vorgestellt, wobei zu berücksichtigen ist: „there is no purely automatic way of identifying phrasal units of meaning“⁵⁶ (Stubbs 2007: 181).

5.2.1 FREQUENZ UND DISPERSION

Für den Spracherwerb ist die Häufigkeit im Sprachgebrauch die wichtigste Einflussgröße. Je häufiger ein Wort oder eine Konstruktion im Input vorkommt, desto wahrscheinlicher ist es, dass Lernende der Konstruktion begegnen. Zur Extraktion von formelhaften Sequenzen werden dafür in der Regel Frequenzlisten von *n*-Grammen, d.h. Mehrworteinheiten, erstellt. Zu beachten ist, dass die laufenden Wortformen eines Textes, die Tokens, unterschiedlich definiert sein können, Beispiele wie *Brandenburger Tor* oder *Lotto-Annahmestelle* zeigen, dass nicht immer eindeutig ist, was als Wort zählt (vgl. Gries/Ellis 2015: 4).

Ferner sagt die Frequenz allein nichts darüber aus, wie gleichmäßig die Konstruktionen über das Korpus verteilt sind, also wie regelmäßig Lernende ihnen begegnen. Dies wird mit dem statistischen Maß *Dispersion DP* (Gries 2008c) erfasst. Dafür wird die Größe einzelner Korpusteile anteilig am gesamten Korpus sowie die Frequenz einer Konstruktion in einem Korpusteil anteilig an allen Vorkommen der Konstruktion im gesamten Korpus gemessen. Ein geringer DP Wert bedeutet eine ausgewogene, ein hoher Wert dementsprechend eine unausgeglichene Verteilung. Das Maß ist insofern relevant, als es im Vergleich zu den Frequenzen besser mit psycholinguistischen Ergebnissen wie Reaktionszeiten zu korrelieren scheint, wobei nicht genügend Daten vorliegen, um wirklich belastbare Aussagen treffen zu können (vgl. Gries 2010: 208).

Ein weiteres Häufigkeitsmaß ist die Type-Frequenz, die, wie bereits in 3.1.2.1 erläutert, bezeichnet, wie viele unterschiedliche lexikalische Elemente die Leerstelle einer Konstruk-

⁵⁶ ‚Bedeutungseinheiten können nicht rein automatisch identifiziert werden.‘ (eigene Übersetzung).

tion füllen können. Um die Verteilung der Typen zu erfassen, kann das Maß *entropy*, das Werte zwischen 0 und 1 erzielt, bemüht werden; je höher der Wert, desto ausgeglichener die Verteilung (vgl. Gries/Ellis 2015: 8).

5.2.2 ASSOZIATIONSMAßE

Assoziationsmaße wie *Pointwise MI (Mutual Information)* geben an, inwieweit Wörter öfter zusammen vorkommen als der Zufall erwarten lässt. Höhere MI-Werte stehen für eine stärkere Anziehungskraft zwischen den Wörtern, niedrigere Werte legen nahe, dass diese eher zufällig zusammen vorkommen. Dabei bevorzugen verschiedene Assoziationsmaße unterschiedliche Mehrworteinheiten. Während bspw. das Maß *t-score* häufige Kollokationen extrahiert, generiert das Maß MI hohe Werte für wenig frequente Wörter (vgl. Durrant/Schmitt 2009: 167). Während klassische Kollokationsanalysen von Lexemen ausgehen, erweitern Stefanowitsch und Gries (2003) diese Methodik um grammatische Konstruktionen, d.h. sie machen die Wechselbeziehungen, die zwischen Lexemen und grammatischen Konstruktionen bestehen, zu ihrem Untersuchungsgegenstand. Dabei bestimmen sie, welche Lexeme häufig bzw. selten in einer bestimmten Leerstelle vorkommen.

Grundsätzlich können drei unterschiedliche Herangehensweisen und Zielsetzungen korpuslinguistischer Untersuchungen unterschieden werden (vgl. Callies 2015: 36)⁵⁷: (i) *corpus-informed* – das Korpus dient lediglich als Referenz, um eine Annahme z.B. zur Frequenz eines Chunks zu überprüfen (ii) *corpus-based* – linguistische Vorannahmen und Theorien werden mittels Korpora veranschaulicht oder überprüft, z.B. können lernersprachliche Daten mit muttersprachlichen Daten verglichen werden und (iii) *corpus-driven* – eine induktive Herangehensweise, die ohne Vorannahmen auskommt und (Ir)Regularitäten bzw. Muster in den Sprachdaten identifiziert. Während (i) nicht mit den Korpusdaten per se arbeitet, dienen sie in (ii) und (iii) als Grundlage für die Analyse. Für den Untersuchungsgegenstand der vorliegenden Arbeit eignen sich die Herangehensweisen *corpus-based* sowie *corpus-driven*; Zum einen sollen rekurrente Muster in den Sprachdaten offengelegt werden, zum anderen sollen Hypothesen bzgl. der Lernersprache im Vergleich zum muttersprachlichen Gebrauch überprüft werden, wie nachfolgend aufgeführt.

⁵⁷ Vgl. hierzu auch Steyers (2013: 77) Unterteilung im Hinblick auf Wortverbindungen: (i) Wortverbindungen im Korpus mittels Phrasensuche verifizieren, (ii) Wortverbindungen im Korpus mittels statistischer Berechnungen identifizieren und (iii) Wortverbindungen im Korpus ohne Vorannahme finden.

6. ANALYSE

6.1 ZIELSETZUNGEN DER ANALYSE

Zunächst geht es darum, festzustellen, welche Chunks von Lernenden auf einem fortgeschrittenen Niveau des Deutschen als Fremdsprache verwendet werden. Dafür sollen in einem ersten Schritt die 350 frequentesten 3- und 4-Gramme des Lernerkorpus sowie des muttersprachlichen Vergleichskorpus untersucht werden. Bei der Analyse soll nach Möglichkeit zwischen festen Wortverbindungen und flexibleren Wortbindungsmustern (Steyer 2013) unterschieden werden. Laut Steyer (vgl. 2013: 27) spielt die übersummativ Bedeutung der Wendungen dabei eine wichtigere Rolle als die Formseite der Konstruktion; Elemente, die Leerstellen in einem Wortbindungsmuster füllen, müssen sich primär in die ganzheitliche Bedeutung der Sequenz einfügen. In dieser Arbeit sollen jedoch sowohl inhaltsseitige als auch formseitige Beziehungen verschiedener Realisationen von Wortbindungsmustern berücksichtigt werden, um diese unterschiedlichen Kategorien struktureller und semantisch-pragmatischer Art zuordnen zu können. Mittels einer Kookkurrenzanalyse sollen in einem zweiten Schritt assoziative Mehrwortverbindungen aufgedeckt werden, die bei der Frequenzanalyse nicht erfasst wurden. Anschließend soll das Korpus nach alternativen Füllern für die Leerstellen der identifizierten Wortbindungsmuster durchsucht werden, sodass diese zukünftig für die Didaktik nutzbar gemacht und im Unterricht als unterschiedliche Realisationen halb schematischer Chunks eingesetzt werden können. Somit soll beiden Ansätzen – dem kommunikativen als auch dem *Chunk-und-Konstruktionen-Ansatz* – Rechnung getragen werden. Zuletzt wird eine kontrastive Analyse der extrahierten L2- und L1-Chunks durchgeführt, um die in der Forschung zu findende These, dass Lerner zwar frequente, jedoch kaum hochassoziative Wendungen nutzen (Simpson-Vlach/Ellis 2009; Ellis et al. 2008; Durrant/Schmitt 2009), zu überprüfen und um Präferenzen und Restriktionen der verwendeten Chunks zu vergleichen.

Zusammenfassend stellen sich also folgende Fragen für die Analyse:

- (i) Welche Merkmale weisen die extrahierten Chunks im Lernerkorpus sowie im Vergleichskorpus auf?
- (ii) Wie werden halb-schematische Wortbindungsmuster realisiert?
- (iii) Inwieweit unterscheidet sich der Gebrauch formelhafter Chunks von Lernenden und Muttersprachlern?

6.2 DATENGRUNDLAGE

In den nächsten zwei Abschnitten werden das fehlerannotierte Lernerkorpus Falko sowie Informationen zu den Teilnehmenden wie Alter, Geschlecht, Muttersprachen und Themenwahl beschrieben.

6.2.1 DAS LERNERKORPUS FALKO

Das **fehlerannotierte Lernerkorpus Falko** (Lüdeling et al. 2008) ist ein online verfügbares Lernerkorpus des Deutschen als Fremdsprache (<https://korpling.german.hu-berlin.de/falko-suche/search.html>), das Texte von fortgeschrittenen Lernenden enthält. Das Korpus umfasst fünf bzw. sechs Subkorpora, die sich in der Schreibaufgabe (Zusammenfassung und argumentatives Essay) sowie der Muttersprache (Deutsch vs. andere L1) unterscheiden. In Tabelle 2 sind jeweils die Anzahl der Texte und der Tokens pro Korpus aufgelistet.

Aufsätze		Texte	Tokens
Lernerkorpus	falkoEssayL2v2.4	248	144.619
	falkoEssayL2WHIGv2.0	195	130.187
muttersprachliches Kontrollkorpus	falkoEssayL1v2.3	95	70.615

Zusammenfassungen		Texte	Tokens
Lernerkorpus	FalkoSummaryL2v1.2	106	40.638
muttersprachliches Kontrollkorpus	FalkoSummaryL1v1.2	57	21.211

Tabelle 2. Subkorpora des Lernerkorpus Falko.

(Falko ANNIS Corpus Search; 20.10.2016)

Das Aufsatzkorpus umfasst argumentative Aufsätze von Lernenden und Muttersprachlern zu folgenden vier Themen:

- Der Feminismus hat den [Interessen der]⁵⁸ Frauen mehr geschadet als genutzt.
- Kriminalität zahlt sich nicht aus.
- Die meisten Universitätsabschlüsse [sind nicht praxisorientiert und]⁵⁹ bereiten die Studenten nicht auf die wirkliche Welt vor. Sie sind deswegen von geringem Wert.
- Die finanzielle Entlohnung eines Menschen sollte dem Beitrag entsprechen, den er/sie für die Gesellschaft geleistet hat.

(Lüdeling et al. 2012: 24)

Die Daten beider Gruppen wurden dabei unter vergleichbaren Bedingungen erhoben (keine Hilfsmittel, 90 Minuten Zeit). Das Lernerkorpus FalkoEssayL2WHIG unterscheidet sich

⁵⁸ Die Auswertung in 6.4 deutet stark darauf hin, dass der Wortlaut der Aufgabenstellung die Tokenfolge *Interessen der* beinhaltete.

⁵⁹ Grünhagen (2015: 98).

nur insofern von den anderen Lernertexten, als dass mehr Metadaten erhoben wurden. Ferner gelten die Lernenden als fortgeschritten, da sie mindestens 60 von 100 Punkten in einem C-Test der Humboldt-Universität zu Berlin erreichten (vgl. ebd.: 23). Das L1-Korpus wurde an Brandenburger und Berliner Gymnasien (12./13. Klasse) erhoben (vgl. Lüdeling et al. 2008: 68).

In der nachfolgenden Analyse wurden lediglich die Essaykorpora berücksichtigt, da sie weitaus mehr Texte umfassen und davon ausgegangen wird, dass die Sprache insofern authentischer ist, als sie sich nicht an einem Vorlagentext, wie das bei den Zusammenfassungen der Fall ist, sondern lediglich an zu diskutierenden Thesen orientiert. Auf die Textsorte und textsortenspezifische Chunks soll hier nicht genauer eingegangen werden, da mit Stein (2007: 225) konstatiert wird, dass „die meisten Phraseme textsortenspezifisch unmarkiert sind“. Insgesamt bilden also 443⁶⁰ Texte mit 274.806 Tokens die Grundlage für die zu untersuchende Lernaltersprache, ein Text besteht demnach durchschnittlich aus ungefähr 620 Tokens. Das muttersprachliche Vergleichskorpus umfasst 95 Texte mit insgesamt 70.615 Tokens, was einem Wert von ungefähr 743 Tokens pro Text entspricht.

6.2.2 METADATEN DER TEILNEHMENDEN

Metadaten zu den Subkorpora können in Form von Excel-Tabellen auf der Internetseite der HU Berlin⁶¹ heruntergeladen werden. Während in ANNIS 195 Texte für das Subkorpus falkoEssayL2WHIGv2.0 angegeben wurden (s. Tabelle 2), werden in der entsprechenden Excel-Tabelle Metadaten zu 196 Texten bzw. 196 Teilnehmenden aufgeführt. Nachfolgend werden die Metadaten zu den Lernerkorpora falkoEssayL2WHIGv2.0 und falkoEssayL2v2.4 weitestgehend zusammen beschrieben, da sie in der Analyse ebenfalls zusammengefasst werden. Von den 444 Lernenden war der überwiegende Teil, mit 309 Teilnehmerinnen weiblich und 128 männlich; sieben machten keine Angabe. Die Datenerhebung fand in den Jahren 2006, 2007, 2008 (falkoEssayL2v2.4) und 2010, 2011 (falkoEssayL2WHIGv2.0) statt. Die Teilnehmenden wurden zwischen 1943 und 1992, die meisten jedoch nach 1980 geboren. Mit 168 Texten wurde das Thema „Studium“ am häufigsten gewählt, gefolgt von „Feminismus“ (123), „Kriminalität“ (87) und zuletzt „Entlohnung“ (66). 232 Teilnehmende, also mehr als die Hälfte, gaben Englisch als Erstsprache bzw. als

⁶⁰ bzw. 444 Texte, s. 6.2.2.

⁶¹ <<https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko>> (01.11.16).

„am besten beherrschte Muttersprache“ (Lüdeling et al. 2012: 10) an, Tabelle 3 zeigt zusätzlich die Verteilung anderer Muttersprachen.

L1	Anzahl	L1	Anzahl
Afrikaans (afr)	8	Luhya (luy)	2
Katalanisch (cat)	2	Nord-Ndebele (nde)	1
Tschechisch (ces)	6	Niederländisch (nld)	7
Maa (cma)	2	Norwegisch (nor)	9
Dänisch (dan)	38	Polnisch (pol)	10
Neugriechisch (ell)	7	Rumänisch (ron)	3
Englisch (eng)	232	Russisch (rus)	14
Finnisch (fin)	5	Slowakisch (slk)	1
Französisch (fra)	19	Nord-Sami (sme)	1
Serbokroatisch (hbs)	1	Spanisch (spa)	4
Hindi (hin)	2	Albanisch (sqi)	1
Ungarisch (hun)	2	Schwedisch (swe)	4
Sichuan-Yi (iii)	1	Tamil (tam)	1
Italienisch (ita)	5	Tatarisch (tat)	1
Japanisch (jpn)	3	Türkisch (tur)	10
Kikuyu (kik)	5	Ukrainisch (ukr)	8
Koreanisch (kor)	2	Usbekisch (uzb)	11
Oshivambo (kua)	1	Vietnamesisch (vie)	2
Litauisch (lit)	1	Chinesisch (zho)	7
Kiluba (lub)	5		

Tabelle 3. Verteilung der am besten beherrschten Muttersprachen.

Darüber hinaus gaben 221 Teilnehmende, also knapp die Hälfte, an, Deutsch sei die Fremdsprache, die sie am besten beherrschten, für 169 Teilnehmende ist Deutsch die Fremdsprache, die sie am zweitbesten beherrschen (vgl. Lüdeling et al 2012: 10). Beim C-Test erzielten die Teilnehmenden zwischen 60 und 99 Punkten, im Schnitt ungefähr 80 Punkte.

Das muttersprachliche Vergleichskorpus wurde in den Jahren 2006 und 2007 erhoben. Der Großteil der Teilnehmenden war weiblich (69 Teilnehmerinnen), lediglich 26 männliche Teilnehmer nahmen an der Erhebung teil. Die Geburtsdaten liegen zwischen 1964 und 1991, wobei die meisten zwischen 1989 und 1991 geboren wurden. Während die meisten Lernenden das Thema „Studium“ wählten, wurde es von den Muttersprachlern lediglich acht Mal gewählt. 46 entschieden sich für das Thema „Kriminalität“, 26 wählten „Entlohnung“ und 15 „Feminismus“.

6.3 DATENERHEBUNG

In einem ersten Schritt wurden die Rohdaten des Falko Korpus' in Form von Excel-Tabellen heruntergeladen⁶². Für die Datenextraktion wurde anschließend das Programm

⁶² < <https://korpling.german.hu-berlin.de/falko-suche/search.html> > (20.09.2016).

falkotool.py geschrieben⁶³, das folgende Funktionen ausführt: (1) Lies Sätze (=Sequenz von (Token, Lemma, POS-Tag) Tupeln) aus Excel-Tabellen; (2) Erstelle n-Gramm-Frequenzlisten aus Sätzen, wobei der Benutzende wählen kann, ob die n-Gramme aus Tokens, Lemmata oder POS-Tags bestehen; (3) Benutze Frequenzlisten, um PMI zwischen n-Grammen zu bestimmen. PMI ist dabei folgendermaßen definiert:

$$PMI(a, b, c) = \ln \frac{p(a, b, c)}{p(a)p(b, c)}$$

PMI steht für *Pointwise Mutual Information* und ist ein Assoziationsmaß (s. 5.2.2), das angibt, wie stark zwei oder mehrere Begriffe statistisch voneinander abhängig sind. Es geht also darum, wie viele Informationen in einer Variablen über eine andere enthalten sind. So werden n-Gramme extrahiert, deren einzelne Komponenten nicht unbedingt häufig, aber wenn, dann zusammen vorkommen. Da für sehr seltene n-Gramme ein zu hoher MI-Wert generiert wird, wurde für die Extraktion ein Frequenzminimum von fünf Vorkommen gesetzt (vgl. Marschner/Langer 2006: 32).

Zunächst wurden also 3- und 4-Gramm-Frequenzlisten für das Lernerkorpus sowie das muttersprachliche Vergleichskorpus erstellt, d.h. die 350 frequentesten kombinatorisch möglichen 3- und 4-Gramme wurden extrahiert. Abbildung 4 zeigt einen Ausschnitt der häufigsten Trigramme mit den jeweiligen Frequenzen im Lernerkorpus⁶⁴.

```
In [5]: ngram_freqs = falkotool.load_ngrams("./freqsn3essay12.json")

In [6]: ngram_freqs[3].most_common(350)

Out[6]: [(u'die wirkliche welt', 299),
          (u'interessen der frauen', 242),
          (u'auf die wirkliche', 238),
          (u'den interessen der', 214),
          (u'von geringem wert', 199),
          (u'f\xfc die gesellschaft', 194),
          (u'nicht auf die', 162),
          (u'die studenten nicht', 157),
          (u'wirkliche welt vor', 152),
          (u'der frauen mehr', 146),
          (u'studenten nicht auf', 145),
          (u'die meisten universit\xe4tsabschl\xfcsse', 143),
          (u'bereiten die studenten', 133),
          (u'meiner meinung nach', 132),
```

Abbildung 4. Ausschnitt der häufigsten Trigramme im Lernerkorpus.

⁶³ An dieser Stelle möchte ich mich sehr herzlich bei Johannes Baiter bedanken, der das Programm geschrieben hat.

⁶⁴ Wie der Ausschnitt zeigt, berücksichtigt das Programm keine Groß- und Kleinschreibung. Dies ist für die Zielsetzung der Arbeit auch nicht relevant. In der nachfolgenden Analyse werden Substantive großgeschrieben.

In einem nächsten Schritt wurden – jeweils für das Lerner- und das Vergleichskorpus – Assoziationslisten erstellt, d.h. Listen, in denen die 3- und 4-Gramme absteigend nach den PMI-Werten geordnet wurden, wie in Abbildung 5 für das Vergleichskorpus veranschaulicht.

```
In [7]: ngram_freqs = falkotool.load_ngrams("./freqsn3essay11.json")

In [8]: falkotool.get_collocations(ngram_freqs, n=3, min_freq=5)[:350]

Out[8]: [(u'voll und ganz', 8.535230607705435),
(u'fr\xfcher oder sp\xe4ter', 8.124488442496427),
(u'art und weise', 7.634940217177205),
(u'geschadet als gen\xftzt', 7.572419860196386),
(u'l die finanzielle', 7.420869962069186),
(u'entsprechen den er', 7.3738174177982945),
(u'finanzielle entlohnung eines', 7.367625447550373),
(u'gro\xdfen und ganzen', 7.355355354958559),
(u'beitrag entsprechen den', 7.117683703081439),
(u'meiner meinung nach', 7.090284728893325),
(u'studenten nicht auf', 6.994785566758285),
(u'entlohnung eines menschen', 6.852885924463246),
(u'zur rechenschaft gezogen', 6.834820917065607),
(u'interessen der frauen', 6.702649143957124),
```

Abbildung 5. Ausschnitt assoziativer Trigramme im Vergleichskorpus.

Abschließend wurden einige auffällige Chunks noch einmal manuell in ANNIS gesucht, um Gebrauchspräferenzen und Restriktionen herauszufiltern. Dafür wurde jeweils eine Zufallsstichprobe von fünf Instanzen untersucht.

6.4 AUSWERTUNG

Von den 350 frequentesten 3- und 4-Grammen wurden zunächst in beiden Korpora die Sequenzen aussortiert, die dem exakten Wortlaut der Aufgabenstellung bzw. der zu diskutierenden Aussagen (s. 6.2.1) entsprachen (z.B. *die wirkliche Welt*, *von geringem Wert*, *die finanzielle Entlohnung*, vgl. auch Abb.4). Anschließend wurden lediglich die n-Gramme berücksichtigt und weiter analysiert, die im Sinne Steyers (vgl. 2013: 24) eine kommunikative Einheit bilden, d.h., jene, die ohne zusätzliche sprachliche Elemente verständlich sind. Dabei wurden jedoch auch erweiterte Chunks angenommen, d.h., Einheiten, welche die Größe von 3- und 4-Grammen überschreiten, wie die denkbare Einheit *[[ich denke dass] [Nebensatz]]*, die sich aus dem Trigramm *ich denke dass* ergibt. Diese entsprechen den gesprächsspezifischen Formeln, wie sie in 2.2 beschrieben wurden. Grundsätzlich muss berücksichtigt werden, dass die Entscheidung, ob etwas als Chunk gilt oder nicht, intuitiv erfolgte und es, abgesehen von dem Status als kommunikative Einheit und dem rekurrenten Auftreten der Sequenz, keine „objektiv systematisierbare[n] Kriterien“ (Hein/Bubenhofer 2015: 203) gibt, die eindeutig für den Chunkstatus einer Sequenz spre-

chen⁶⁵. Anhand des Kriteriums „kommunikative Einheit“ wurden folglich falsch positive Treffer wie *der Feminismus den* oder *ist es nicht* von der weiteren Analyse ausgeschlossen. Darüber hinaus wurden ternäre und quaternäre Sequenzen ausgeschlossen, die eine Mindestfrequenz von fünf unterschritten, da sie das Kriterium der Rekurrenz nicht ausreichend erfüllen, sie einen zu hohen MI-Wert generieren und davon ausgegangen wird, dass weniger häufig vorkommende Einheiten zu wenig Aussagekraft haben, um in die Analyse aufgenommen zu werden. Es muss jedoch festgehalten werden, dass damit nicht etwa ein Schwellenwert für eine kognitive Verankerung von Chunks festgelegt werden soll, denn „[t]he difference between higher and lower frequency is one of degree (the level of activation), and not kind (stored vs. computed)“⁶⁶ (Arnon/Snider 2010: 69).

6.4.1 EIGENSCHAFTEN VON L2-CHUNKS

Bei der Kategorisierung der n-Gramme wurde klar, dass es unterschiedliche Gründe für ihr Auftreten gibt. Einige sind thematisch bedingt (z.B. *die Rechte der Frauen*), andere sind gesprächsspezifisch (z.B. *ich denke dass; und so weiter*), wieder andere stellen grammatische Beziehungen her und übernehmen bspw. die Aufgabe von Präpositionen (z.B. *in Bezug auf, im Vergleich zu*) (vgl. Forgács 2009: 132). Tabellen 4 und 5 zeigen eine Auswahl unterschiedlicher Arten drei- und viergliedriger Chunks sowie deren Merkmale und die absoluten und relativen Frequenzen.

Chunk	Freq.	Freq. pro Mio. ⁶⁷	Merkmal	Kriterium
meiner Meinung nach	132	480	gesprächsspezifische Formel, meinnungsäußernd	pragmatisch
an der Universität	115	418	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
die Mehrheit der	21	76	Nominalphrase, Genitivattribut	strukturell
im Vergleich zu	21	76	Präpositionalphrase, strukturelles Phrasem	strukturell
die gleiche Rechte	14	50	fehlerhaft	strukturell
ab und zu	12	44	Paarformel	strukturell

Tabelle 4. Auswahl dreigliedriger Chunks.

⁶⁵ Im Anhang finden sich die Listen 1-8 mit allen extrahierten 3- und 4-Grammen, sodass eventuell andere Sequenzen, die ich ausgeschlossen habe, untersucht werden können.

⁶⁶ „Der Unterschied zwischen häufigem und weniger häufigem Vorkommen ist ein gradueller (Grad an Zugänglichkeit) und kein kategorisierender (ganzheitlich gespeichert vs. neu generiert)“ (eigene Übersetzung).

⁶⁷ Für die Berechnung der relativen Frequenzen wurde eine Gesamttokenanzahl von 274.806 Tokens angenommen, wie sie in ANNIS für die Summe von falkoEssayL2v2.4 + falkoEssayL2WHIGv2.0 angegeben ist. Zusätzlich ist zu beachten, dass die maximal mögliche relative Häufigkeit für ein Trigramm 333.333 und für ein 4-Gramm 250.000 beträgt.

Chunk	Freq.	Freq. pro Mio.	Merkmal	Kriterium
auf der anderen Seite	47	171	Präpositionalphrase, strukturelles Phrasem	strukturell
in der heutigen Gesellschaft	13	47	Präpositionalphrase, lokal, nicht zielgerichtet, übertragener Gebrauch	strukturell
es ist klar dass	13	47	gesprächsspezifische Formel, evaluativ	pragmatisch
auf der andere Seite	7	25	fehlerhaft	strukturell

Tabelle 5. Auswahl viergliedriger Chunks.

6.4.1.1 DREIGLIEDRIGE CHUNKS⁶⁸

Für das Lernerkorpus wurden insgesamt 76 dreigliedrige Chunks identifiziert. Mit 132 Vorkommen tritt der Chunk *meiner Meinung nach* am häufigsten auf, die niedrigste Frequenz eines Chunks – innerhalb der 350 untersuchten Trigramme – liegt bei zwölf, der Median bei 18. Die größte Gruppe bilden Präpositionalphrasen mit 31 Vorkommen (vgl. für die nähere semantische Bestimmung Helbig/Buscha 2001: 362-390). Davon bezeichnen zwölf einen Ort und sind nicht zielgerichtet (z.B. *an der Universität* (96)⁶⁹; *in der Schule* (30); *in den USA* (21)). Vier werden in einem übertragenen Sinn gebraucht: *in der Gesellschaft* (115); *in der Arbeitswelt* (33); *in unserer Gesellschaft* (25); *in der Praxis* (20). Sieben Präpositionalphrasen sind temporal (z.B. *in der Vergangenheit* (32); *in der Zukunft* (27); *nach dem Abschluss* (13); *im Laufe der X* (12)), eine kausal (*wegen des Feminismus* (13)) und eine Phrase wird modal gebraucht (*für die Arbeitswelt* (12)). Die genannten Chunks sind alle relativ flexibel, zählen also eher zu den Wortverbindungsmustern. Unter den Präpositionalphrasen finden sich jedoch auch fünf relativ feste Wortverbindungen: *auf jeden Fall* (35); *in Bezug auf* (29); *in diesem Fall* (24); *im Vergleich zu* (21); *im Vergleich mit* (15). Drei davon zählen zu den strukturellen Phrasemen (*in Bezug auf* (29); *im Vergleich zu* (21); *im Vergleich mit* (15)).

Dabei fällt es sehr schwer, eine Grenze zwischen festen und flexiblen Wortverbindungen zu ziehen, diese wird eher durch einen fließenden Übergang charakterisiert. So sind die Chunks *im Vergleich zu* und *im Vergleich mit* nur sehr eingeschränkt veränderbar, nichtsdestoweniger kann ein lexikalisches Element ausgetauscht werden, sodass sie laut Steyer (2013) zu den Wortverbindungsmustern zählen. Darüber hinaus stellt sich die Frage, wie abstrakt die übersummativ Bedeutung einer Usuellen Wortverbindung sein darf, also ob z.B. *in der Schule* und *in der Gesellschaft* demselben Wortverbindungsmuster angehören. Die unterschiedlichen Bedeutungsdimensionen, die, je mehr Chunks dazugezählt werden, immer abstrakter werden, sind in Abbildung 6 veranschaulicht.

⁶⁸ Vgl. Tabelle 1. L2_n3_mostcommon im Anhang

⁶⁹ Die absoluten Frequenzen der Chunks werden jeweils in Klammern angegeben.

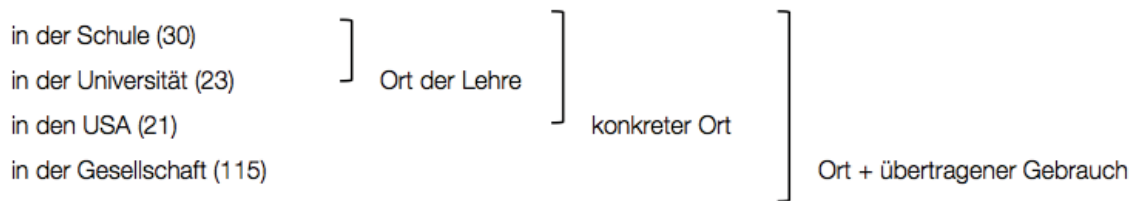


Abbildung 6. Bedeutungsdimensionen des Schemas *in d X*⁷⁰.

Ferner lässt sich in Frage stellen, ob lediglich Füller zugelassen werden, die nur eine geringfügige Bedeutungsänderung herbeiführen und so die übersummativ Funktion des Wortverbindungsmusters erhalten, d.h. quasi synonyme Realisationen eines Musters oder auch Füller, die andere semantische Relationen wie Antonymie darstellen. Mit anderen Worten, gilt *in der Praxis* (20) als feste Wortverbindung, dessen lexikalische Komponenten nicht ausgetauscht werden können oder als Wortverbindungsmuster, das semantisch mit der Sequenz *in der Theorie* verbunden ist?

Neben den Präpositionalphrasen wurden zwölf Nominalphrasen identifiziert. Sechs davon beinhalten ein Adjektivattribut (z.B. *ein großes Problem* (18); *die neuen Studiengänge* (13); *das neue System* (12)) und vier ein Genitivattribut mit jeweils einer Leerstelle (*die Mehrheit der X* (21); *am Ende des X* (19); *die Rolle der X* (15); *das Leben der X* (15)). Darüber hinaus findet sich eine Genitiversatzform (*die Mehrheit von X* (15)), die in 6.4.2 noch genauer untersucht wird. Schließlich lässt sich noch die objektsprachliche Nominalphrase *das Wort Feminismus* (12) identifizieren.

Darüber hinaus wurden die Paarformel *ab und zu* (12) und das flexible Wortverbindungsmuster *Männer und Frauen* (43) bzw. *Frauen und Männer* (34) extrahiert. Hier würde ich intuitiv die Reihenfolge *Männer und Frauen* bevorzugen, eine Restriktion, die für den Chunkstatus der Sequenz spräche. Der Chunk kommt allerdings nicht im muttersprachlichen Vergleichskorpus vor, sodass eine Überprüfung dieser Vermutung aussteht⁷¹.

Folgende identifizierte Lernerchunks weisen formseitig Fehler auf: *im Bezug auf* (17), *meine Meinung nach* (16), *die gleiche Rechte* (14), *die meisten Universitätsabschlüsse* (14) und *von geringen Wert* (13).

⁷⁰ Nach Steyer (2013) liegt dem Schema *in d X* das Strukturmuster *in d NP* mit den NP-Füllern *Schule, Universität, USA, Gesellschaft* zu Grunde.

⁷¹ Zudem kann angenommen werden, dass Lerner diesen Chunk in den Essays zum Thema „Feminismus“ verwendet haben, sodass sie sich evtl. bewusst dafür entschieden haben, Frauen zuerst zu nennen.

Nachdem sich die bis hierhin behandelten Chunks formal kategorisieren lassen, eignen sich die nachfolgenden Chunks für eine pragmatische Analyse, wobei sie i.d.R. die rekurrente Funktion erfüllen, inhaltlich Komplexes einzubetten und in diesem Zusammenhang keine selbstständigen Äußerungseinheiten darstellen (vgl. Filatkina 2007: 145). Elf werden genutzt, um ein Beispiel zu geben, die meisten beinhalten die unpersönliche Form *es gibt* (z.B. *wie zum Beispiel* (25); *zum Beispiel wenn* (19); *es gibt noch* (14); *heutzutage gibt es* (12)). Drei Chunks dienen der Meinungsäußerung (*meiner Meinung nach* (132); *ich glaube dass* (45); *ich denke dass* (18)), drei haben eine thematisierende Funktion (*die Frage ist* (18); *die Frage ob* (14); *die Tatsache dass* (12)) und ein Chunk wird argumentativ gebraucht (*deshalb ist es* (16)). Darüber hinaus erfüllen die Chunks *es ist klar* (14) und *das Problem ist* (13) eine evaluative Funktion – dazu zählen eventuell auch folgende Chunks: *aber es ist* (31) bzw. *es ist aber* (18), wobei diese genauer untersucht werden müssen, um zu klären, ob ihnen ein evaluatives Adjektiv folgt. Schließlich wurden die relativierenden Sequenzen *man könnte sagen* (16) und *es kann sein* (12) sowie die relativ sinnentleerte Einheit *und so weiter* (30) gefunden. Genauso wie die strukturelle Einordnung, lässt sich auch die Funktion eines Chunks nicht immer eindeutig bestimmen, so erfüllt die Sequenz *das Problem ist* neben der evaluativen auch eine thematisierende Funktion, m.a.W. Chunks erfüllen in der Regel mehrere Funktionen, wobei ihnen jeweils die Funktion zugeordnet wurde, die m.E. dominiert. Als einzige Kollokation wurde die Sequenz *zu Hause bleiben* (28) identifiziert (vgl. Häcki Buhofer et al. 2014: 376).

6.4.1.2 VIERGLIEDRIGE CHUNKS⁷²

Die Frequenzen der untersuchten Vierworteinheiten liegen zwischen fünf, was als Minimalfrequenz gesetzt wurde und 47 (*auf der anderen Seite*), der Median liegt bei 7,5. Oftmals stellen die Chunks eine Erweiterung bzw. Spezifizierung der bereits identifizierten dreigliedrigen Chunks dar (z.B. *in der wirklichen Welt* (42); *man könnte sagen dass* (14); *das Leben der Frauen* (9); *die Rolle der Frauen* (7); *in Deutschland gibt es* (7)). Die Tatsache, dass das erweiterte Trigramm weniger häufig als das Trigramm an sich vorkommt, deutet auf ein abstraktes, produktives Muster mit weiteren Füllern hin. So kommt bspw. der Chunk *die Rolle der* 15 Mal, im Vergleich dazu *die Rolle der Frauen* jedoch nur siebenmal vor.

⁷² Vgl. Tabelle 2. L2_n4_mostcommon im Anhang

Unter den nicht erweiterten Chunks⁷³ finden sich drei strukturelle Phraseme (*auf der anderen Seite* (47); *auf der einen Seite* (11); *es ist nicht nur* (8)), wobei geklärt werden muss, ob *es ist nicht nur* tatsächlich dem strukturellen Phrasem *nicht nur ... sondern/auch* entspricht. Ferner lassen sich vier Präpositionalphrasen (temporal – *in den letzten Jahren* (14); lokal, zielgerichtet – *in die wirkliche Welt* (13); übertragener Gebrauch – *zwischen Männern und Frauen*⁷⁴ (12); *in den meisten Fällen* (7)) und drei Nominalphrasen mit Genitivattributen (*die Interessen* (24) – ; *die Rechte* (9) – ; *die Situation* (5) – *der Frauen*). Darüber hinaus wurde die Paarformel *im Großen und Ganzen* (9) sowie der Finalsatz *um Geld zu verdienen* (8) extrahiert, der außerdem die Kollokation *Geld verdienen* enthält (vgl. Häcki Buhofer et al. 2014: 316).

Vier Chunks wurden formseitig fehlerhaft produziert (*zwischen Männer und Frauen* (11); *die Interesse der Frauen* (10); *auf der andere Seite* (9); *in der wirkliche Welt* (7)).

Die gesprächsspezifischen Chunks lassen sich analog zu den dreigliedrigen Chunks einteilen in meinungsäußernd (*ich bin der Meinung* (17); *ich würde sagen dass* (7)), ein Beispiel gebend (*wenn man z. B.* (6); *wenn man zum Beispiel* (6)), relativierend (*das bedeutet nicht dass* (7); *es kommt darauf an* (6); *es ist möglich dass* (6)), thematisierend (*man darf nicht vergessen* (5)) und argumentativ (*es gibt viele Gründe* (5)). Zudem lässt sich das textstrukturierende Phrasem *zum Schluss möchte ich* (6) identifizieren. Neben *zum Schluss möchte ich* wurde ebenfalls der Chunk *zum Schluss glaube ich* (7) gefunden, der in 6.4.2 hinsichtlich seiner Verwendung noch genauer analysiert wird.

6.4.1.3 ASSOZIATIVE CHUNKS⁷⁵

Die Assoziationsanalyse hat für die dreigliedrigen Chunks, neben dem Eigennamen *Simone de Beauvoir*, auffällig viele Paarformeln (*Vor- und Nachteile*; *Art und Weise*; *Vorteile und Nachteile*; *alles in allem*; *mehr oder weniger*) extrahiert. Ferner wurden elf Präpositionalphrasen (z.B. temporal – *seit vielen Jahren*, *vor 50 Jahren*; lokal, volkswirtschaftlicher Bereich – *bei einer Firma*; partitiv – *mit den Kindern*) identifiziert. Unter den neun gefundenen Nominalphrasen enthalten vier ein Adjektivattribut (z.B. *eine gute Idee*; komparativ – *ein besseres Leben*). Außerdem kommt das Wortverbindungs muster *immer mehr X* auf-

⁷³ Diese sind in der Tabelle 2. L2_n4_mostcommon blau markiert.

⁷⁴ Auch hier werden Männer zuerst genannt, vgl. 6.4.1.1.

⁷⁵ Vgl. Tabelle 3. L2_n3_PMI und Tabelle 4. L2_n4_PMI im Anhang. Es wurden lediglich die Chunks, die nicht schon bei der einfachen Frequenzanalyse extrahiert wurden, berücksichtigt. Diese sind in den Tabellen blau markiert.

fällig oft vor (dreimal: *immer mehr* – *Studenten*; *Leute*; *Frauen*). Ferner wurden die Ersatzform des Genitivs *Typen von Kriminalität*, die auf das Muster *Typen von X* hindeutet, sowie die Temporalbestimmung *den ganzen Tag* identifiziert. Das extrahierte Trigramm *egal ob man* lässt den Chunk bzw. das strukturelle Phrasem *ob ... oder (ob)* vermuten.

Die Chunks *zwischen den Geschlechtern*, *auf jedem Fall* und *an der Universitäten* weisen formseitig Fehler auf. Der Chunk *ich meine dass* wird in 6.4.2 noch einmal genauer untersucht.

Darüber hinaus wurden acht gesprächsspezifische Chunks identifiziert, welche die Funktionen *erklären* (*d. h. dass*; *dies bedeutet dass*), *evaluieren* (*am wichtigsten ist*; *ich finde es*), *argumentieren* (*aus diesem Grund*), *strukturieren* (*wie schon erwähnt*) und *Meinung äußern* (*ich hoffe dass*; *ich finde dass*) erfüllen, wobei sich die Funktionen *evaluieren* und *Meinung äußern* sehr ähneln.

Die Assoziationsanalyse für 4-Gramme ergab lediglich sechs noch nicht analysierte Chunks, die teilweise eine Spezifizierung eines dreigliedrigen Chunks darstellen und sich folgendermaßen kategorisieren lassen:

Chunk	PMI	Merkmal	Kriterium
überall in der Welt	6,8	Adverb + Präpositionalphrase	strukturell
an der Wirtschaftsuniversität Kopenhagen	5,8	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
könnte gesagt werden dass	5,7	gesprächsspezifische Formel, relativierend	pragmatisch
an der Universität studieren	5,6	Präpositionalphrase, Kollokation	strukturell, semantisch
für mich ist es	4,8	gesprächsspezifische Formel, evaluativ	pragmatisch
Frauen in der Politik	4	Nominalphrase	strukturell

Tabelle 6. Viergliedrige assoziative Chunks.

Zusammenfassend sollen nun die drei produktiven Schemata *die Mehrheit der X*, *in der X* und *immer mehr X* aufgeführt werden. Die jeweiligen Sequenzen wurden dafür via ANNIS in den Lernerkorpora *falkoEssayL2WHIGv2.0* und *falkoEssayL2v2.4* gesucht. Anschließend wurde eine zufällig ausgewählte Stichprobe von fünf Instanzen, wie beispielhaft in Tabelle 7 aufgezeigt, analysiert.

1) Path: <i>falkoEssayL2WHIGv2.0</i> > BNG2-2010-11-121 (tokens 182 - 195)	die Mehrheit der Frauen
2) Path: <i>falkoEssayL2v2.4</i> > fkb042_2008_08_L2v2.4 (tokens 1 - 9)	die Mehrheit der Universitäten
3) Path: <i>falkoEssayL2v2.4</i> > fkb011_2008_07_L2v2.4 (tokens 363 - 376)	die Mehrheit der Sachen
4) Path: <i>falkoEssayL2WHIGv2.0</i> > BNG-2010-03-025 (tokens 256 - 269)	die Mehrheit der Uni-Studenten

5) Path: FalkoEssayL2WHIGv2.0 > BNG2-2011-03-204
(tokens 568 - 581)

die Mehrheit der Bevölkerung

Tabelle 7. Chunk / [Dd]je/ . / [Mm]ehrheit/ . "der" . pos="NN" (16)⁷⁶.

Aus den untersuchten Vorkommen ergibt sich folgendes Schema:

- (6) Schema: *die Mehrheit der X*
 Strukturmuster: *die Mehrheit der N-P*
 N-P-Füller: *Frauen/(Uni-)Studenten/Bevölkerung*

Dabei scheint das Schema semantische Füller zu bevorzugen, die eine Gruppe, besonders von Menschen, bzw. das weibliche Kollektivum *Bevölkerung*, das ebenfalls eine Personengruppe darstellt, bezeichnen, denn *Universitäten* ist m.E. nur bedingt zulässig und *Sachen* ist m.E. kein akzeptabler Füller für X⁷⁷.

Für die Chunks *in der X* und *immer mehr X* ergeben sich folgende Schemata:

- (7) Schema: *in der X*
 Strukturmuster: *in der N-P*
 N-P-Füller⁷⁸: *Gesellschaft/Praxis/Grundschule/Welt*
- (8) Schema: *immer mehr X*
 Strukturmuster: *immer mehr N-P*
 N-P-Füller⁷⁹: *Leute/Geld/Kritik/Menschen/Jugendliche*

6.4.1.4 ERWEITERTE L2-CHUNKS

Einige Äußerungseinheiten – *meiner Meinung nach*; *ich glaube dass*; *aber es ist* bzw. *es ist aber*; *für mich ist es*; *es ist nicht nur*; *egal ob man*; *es kommt darauf an* – sollen nun noch einmal genauer hinsichtlich ihrer Eigenschaften untersucht werden. Dafür wurden die Chunks ebenfalls in ANNIS gesucht, wonach eine Zufallsstichprobe von fünf Beispielen analysiert wurde. Anhand der ausgewählten Sequenzen soll überprüft werden, inwieweit die Lernenden erweiterte Chunks, die abstrakte Elemente enthalten, (korrekt) verwenden⁸⁰.

Die untersuchten Instanzen legen nahe, dass für den Chunk *meiner Meinung nach* ein erweiterter, lexikalisch offener Chunk [[*meiner Meinung nach*] [VFIN]] angenommen werden

⁷⁶ Die Frequenzen, die in ANNIS angezeigt werden, unterscheiden sich geringfügig von den Frequenzen, die das Programm errechnet hat. Dies lässt sich mitunter auf die Zeichensetzung zurückführen. Die Analysen der Chunks "in" . "der" . pos="NN" (517) und / [Ii]mmer/ . "mehr" . pos="NN" (44) sind im Anhang unter (a) Wortverbindungsmuster aufgeführt.

⁷⁷ Angesichts der Schwierigkeit „alle – wäre das jemals möglich? – form- und inhaltsseitig relevanten Aspekte von Konstruktionen“ (Ziem/Lasch 2013: 140) zu erfassen, beschränke ich mich hier auf die Füller, die tatsächlich von den Lernenden gebraucht wurden.

⁷⁸ Formal füllen lediglich Feminina die N-P Leerstelle; semantisch bezeichnen die Füller einen mehr oder weniger abstrakten Ort.

⁷⁹ Hier sind nur Füller mit Nullartikel zulässig.

⁸⁰ Die analysierten Chunks sind im Anhang unter (b) Erweiterte L2-Chunks aufgeführt.

kann. Auch der Chunk *ich glaube dass* wurde in den analysierten Beispielen korrekt mit Verbendstellung produziert, lässt sich also zu *[[ich glaube dass] [Nebensatz]]* erweitern. Den Chunks *aber es ist* und *es ist aber*, die das unpersönliche *es* enthalten, folgt in acht von den zehn untersuchten Beispielen ein qualifizierendes Adjektiv, sodass sich folgendes Schema ergibt:

(9) Schema: *aber es ist/es ist aber X*

Strukturmuster: *aber es ist/es ist aber* ADJ-P
 ADJ-P-Füller⁸¹: *(auch) nötig/(trotzdem), (auch) wichtig/notwendig/wahr/klar*

Dem viergliedrigen Chunk *für mich ist es* folgt in vier von fünf analysierten Fällen ebenfalls ein evaluatives Adjektiv. In einem Fall folgt ein Adverb: *Für mich ist es hier, dass wir unsere Problem finden*. Die Lernende gab Englisch als am besten beherrschte Muttersprache an; Tatsächlich lässt sich der fehlerhafte Gebrauch auf negativen Transfer aus dem Englischen zurückführen, da folgender Satz akzeptabel ist: *For me, it's here that we find the problem*. Daraus ergibt sich folgendes Schema einschließlich negativer Evidenz :

(10) Schema: *für mich ist es X*

Strukturmuster: *für mich ist es* ADJ-P
 ADJ-P-Füller⁸²: *wichtig/klar/nicht schwierig*
 *Strukturmuster: **für mich ist es* ADV
 *ADV-Füller: *hier*

Für den strukturellen Chunk *es ist nicht nur* wurden zwei Beispiele gefunden, welche die Erweiterung *sondern auch* enthalten. Darüber hinaus wurde ein Beispiel untersucht, dass ebenfalls als Interferenz der L1-Sprache (in diesem Fall ebenfalls Englisch), gedeutet werden kann: *Es ist nicht nur die Kriminelle die leiden, aber auch die Gesellschaft* – *It is not only the criminals who suffer but also society*. So kann auch der fehlerhafte Chunk *nicht nur ... aber auch* als negative Evidenz dienen. Der Chunk *egal ob man* wird von den Lernenden, wie angenommen, als Teil des strukturellen Chunks *ob ... oder* realisiert. Die Analyse des Chunks *es kommt darauf an* ergibt, dass Lernende den Chunk als nicht flexiblen Chunk lernen und nicht in der Lage sind, ihn korrekt in andere Konstruktionen wie einem dass-Nebensatz einzubetten.

⁸¹ Der Füller *am Ende immer wertvoll* wurde nicht aufgeführt, da er nicht so prototypisch und allgemein einsetzbar ist wie die anderen Füller.

⁸² Der Füller *aber schon feministisch* wurde nicht aufgeführt, da er sehr themenspezifisch und m.E. nur bedingt akzeptabel ist.

6.4.2 VERGLEICH L1 – L2

In diesem Abschnitt werden für die häufigsten und assoziativsten L1-Chunks die entsprechenden Frequenzen im Lernerkorpus berechnet, wie in Tabellen acht und neun aufgezeigt⁸³. Damit soll festgestellt werden, ob Lernende diese Sequenzen im Vergleich zu Muttersprachlern zu häufig (overuse) oder zu selten (underuse) gebrauchen. Auch die extrahierten Lernerchunks *Typen von X*; *zum Schluss glaube ich* und *die Mehrheit von*⁸⁴ sollen noch einmal vergleichend für L2 und L1 untersucht werden, um Gebrauchsunterschiede offenzulegen.

Chunk	L1		L2	
	Freq.	Freq. pro Mio.	Freq.	Freq. pro Mio.
meiner meinung nach	36	510	132	480
ich denke dass	14	198	18	66
in unserer gesellschaft	14	198	25	91
in der gesellschaft	14	198	115	418
wie zum beispiel	12	170	25	91
die frage ob	11	156	14	51
in der lage	10	142	7	25
auf der welt	8	113	11	40
der reiche mann	7	99	0	0
auf keinen fall	7	99	11	40

Tabelle 8. Die frequentesten dreigliedrigen L1-Chunks im Vergleich zu L2.

Chunk	L1		L2	
	Freq.	Freq. pro Mio.	Freq.	Freq. pro Mio.
voll und ganz	5	71	0	0
früher oder später	5	71	?	?
art und weise	5	71	10	36
meiner meinung nach	36	510	132	480
gegen das gesetz	6	85	4	15
auf jeden fall	7	99	35	127
ich denke dass	14	198	18	66
was ist mit	6	85	4	15
ich finde es	5	71	9	33
auf keinen fall	7	99	11	40

Tabelle 9. Die assoziativsten dreigliedrigen L1-Chunks im Vergleich zu L2.

Es fällt auf, dass Lernende strukturierende Chunks, die auf Grund ihrer Funktion oftmals im Unterricht verwendet werden wie *meiner Meinung nach* und *auf der anderen Seite* ähnlich häufig oder sogar häufiger als Muttersprachler gebrauchen. Assoziative Chunks wie *voll und ganz*, *Art und Weise*, *im Großen und Ganzen*, *in den meisten Fällen*, *gegen das Gesetz* und *ich denke dass* werden dagegen zu selten genutzt. Während Muttersprachler den Chunk *ich denke dass* (L1: 198 vs. L2: 66)⁸⁵ präferieren, machen Lernende zu häufig Gebrauch von den Wendungen *ich glaube dass* (L1: 57 vs. L2: 164) und *ich meine dass* (L1: 0 vs. L2: 33).

⁸³ Die frequentesten und assoziativsten viergliedrigen L1-Chunks im Vergleich zu L2 sind im Anhang unter (c) Vergleich L1-L2 aufgeführt.

⁸⁴ Die Analysen der Chunks sind im Anhang unter (d) Fehlerhafte Verwendung? aufgeführt.

⁸⁵ Zur besseren Vergleichbarkeit werden hier die relativen Frequenzen angegeben.

Außerdem zeigt sich, dass Lernende den Chunk *in der Gesellschaft* auffällig oft verwenden und diesen im Vergleich zum muttersprachlichen Korpus zu selten variieren: *in der Gesellschaft* (L1: 198 vs. L2: 418), *in unserer Gesellschaft* (L1: 198 vs. L2: 91), *in der heutigen Gesellschaft* (L1: 99 vs. L2: 47).

Für den Chunk *Typen von X* lag die Vermutung nahe, dass er zu häufig gebraucht wird, was womöglich auf Interferenz aus dem Englischen zurückzuführen ist. Die Daten reichen jedoch nicht aus, um eine belastbare Aussage treffen zu können. Für das Lernerkorpus wurden in ANNIS sowohl neun Beispiele für die Sequenz */[Tt]ypen/ . "von" . pos="NN"* als auch für die Sequenz */[Aa]rten/ . "von" . pos="NN"* gefunden. Die Suche im L1-Korpus erzielte dagegen lediglich einen Treffer für */[Aa]rten/ . "von" . pos="NN"* und keinen Treffer für */[Tt]ypen/ . "von" . pos="NN"*. Die Sequenzen *zum Schluss glaube ich* und *die Mehrheit von* mit den absoluten Frequenzen 7 und 15 wurden keinmal von den Muttersprachlern verwendet, was auf zu häufigen Gebrauch seitens der Lernenden hinweist. Dabei ist die Erweiterung *zum Schluss glaube ich dass*, wie sie von den Lernenden verwendet wurde m.E. nicht möglich, da *zum Schluss* eine Reihenfolge, d.h. ein Nacheinander von Aktionen impliziert, *glauben* jedoch einen Zustand beschreibt. Die fünf untersuchten Beispiele, die alle von englischen Muttersprachlern stammen, legen außerdem nahe, dass die Lernenden *zum Schluss* im Sinne von *In conclusion*, ... verwendeten, d.h., um eine Schlussfolgerung aus den präsentierten Argumenten zu ziehen. Die Nominalphrase *die Mehrheit von* nutzen Lernende fehlerhaft als Ersatzform des Genitivs, z.B. in *die Mehrheit von Leuten*⁸⁶.

Zuletzt wurden sowohl für das L1- als auch für das L2-Korpus die Sequenzen *an der Uni* und *und so weiter* identifiziert, wobei *an der Uni* m.E. ein zu umgangssprachliches Register darstellt und *und so weiter* auf Grund der semantischen Leere ebenfalls unpassend scheint. *An der Uni* wurde dabei auffallend oft von L2-Lernenden gebraucht (L1: 42 vs. L2: 197), *und so weiter* (L1: 85 vs. L2: 109) wurde von beiden Gruppen ähnlich häufig realisiert.

⁸⁶ Für Fälle, in denen die Ersatzform des Genitivs mit *von* zulässig oder regelmäßig ist, vgl. Helbig/Buscha 2001: 500f.

6.5 DISKUSSION

Die Auswertung ergab, dass Lernende hauptsächlich Präpositionalphrasen als Chunks speichern, wobei, wie bereits erwähnt, keine endgültige Aussage darüber getroffen werden kann, inwieweit die Sequenzen tatsächlich ganzheitlich abgerufen werden, auch wenn die Token-Frequenz als Indiz für eine hohe kognitive Verankerung gilt. Für den gesteuerten Zweitspracherwerb erscheint es demnach sinnvoll, Präpositionalphrasen als einheitenbildende Chunks, die mit der Zeit variiert werden, zu lernen, um so ein implizites Regelwissen aufzubauen. Dies ist insofern relevant, als „die Flexionsmorphologie [ganz allgemein häufig] als das größte Lernproblem bezeichnet“ wird (Wegener 2010: 206). Ergebnisse der DiGS-Studie (= Deutsch an Genfer Schulen; Diehl et al. 2000) zeigen ebenfalls, dass Präpositionalphrasen chunkartig erworben werden, denn korrekte Kasusmarkierungen in Präpositionalphrasen wurden früher und öfter als in Nominalphrasen realisiert (vgl. ebd.: 299f.). Aguado (2008: 58) schließt daraus, „dass der Zeitpunkt ihres Erwerbs eng mit der kommunikativen Relevanz der jeweiligen Konstruktion verknüpft ist: Präpositionalphrasen [...] werden sehr früh benötigt und deshalb auch früh [...] erworben“. Demnach sollte sich die Auswahl grammatischer Strukturen für den Unterricht an den jeweiligen Kommunikationsbedürfnissen orientieren.

Bezüglich der Formseite von Chunks konnten zudem Forschungsergebnisse, die ergaben, dass Lernende den Fokus auf den Inhalt und nicht auf formseitige Charakteristika legen, für fortgeschrittene Lernende bestätigt werden. Einige feste Wendungen wurden sehr ähnlich, aber formseitig fehlerhaft produziert. Darunter fanden sich u.a. die Wortfolgen *die meisten Universitätsabschlüsse* und *von geringen Wert*, die beide der Aufgabenstellung entnommen wurden. Dies spricht für die Annahme, dass gering saliente, grammatische Informationen von den Lernenden nicht wahrgenommen werden.

Es konnten außerdem einige produktive Wortverbindungsmuster herausgearbeitet werden. Da die Suche auf Lexeme beschränkt war und bspw. keine ternären bzw. quaternären POS-Einheiten gesucht wurden, befinden sich die Schemata eher am lexikalisch spezifischen Ende des Lexikon-Syntax-Kontinuums. Infolgedessen konnte auch kein Prototyp für eine abstrakte Konstruktion ermittelt werden, wie Goldberg et al. (vgl. 2004: 299) dies bspw. für die *Intransitive-Motion*-Konstruktion mit dem prototypischen Verb *go* erarbeiteten und was lernfördernd im Unterricht eingesetzt werden könnte.

Insgesamt bietet die Analyse von Lernerdaten den Vorteil, dass aufgedeckt wird, was Lernende tatsächlich (in)korrekt verwenden. So müssen m.E. nicht alle Restriktionen für eine Konstruktion im Unterricht behandelt werden. Lehrende können sich darauf konzentrieren, was Lernende tatsächlich falsch machen und dies als negative Evidenz nutzen⁸⁷. Die Fehler ließen sich in vielen Fällen auf die Muttersprache zurückführen, sodass kontrastive Analysen neue Erkenntnisse liefern können. Der Vergleich mit dem L1-Korpus bestätigte darüber hinaus, dass Lernende tatsächlich zu wenig Gebrauch von einigen assoziativen Chunks machen. Ferner wurden einige pragmatisch nicht bzw. nur eingeschränkt angemessene Verwendungsweisen aufgedeckt, was eine Chance für eine Vermittlung darstellt, die neben formseitigen Aspekten auch Dimensionen der Bedeutungsseite von Chunks beachtet. Während Lehrwerke bspw. meist suggerieren, dass *ich denke/glaube/meine, dass* gleichwertig verwendet werden können, zeigte die Korpusanalyse, dass Muttersprachler die Sequenz *ich denke, dass* bevorzugen, *ich glaube, dass* nicht ganz so häufig und *ich meine dass* (in dem untersuchten Korpus) nie verwenden.

Zuletzt konnte gezeigt werden, dass die Mehrzahl der Lernenden, im Gegensatz zu Muttersprachlern, den Chunk *in der Gesellschaft*, nicht variierten, was darauf hindeutet, dass einfache Chunks vor komplexen gelernt werden, wobei, wie eben erwähnt, auch produktive Muster in den Lernerdaten gefunden wurden. Auch der Chunk *es kommt darauf an* wurde als lexikalisch feste Einheit gelernt, was zu fehlerhaftem Gebrauch führt. So scheint es sinnvoller, stattdessen drei Chunks zu vermitteln: (i) *[[es kommt darauf an][indirekte Frage]]*, (ii) *[[ich denke, dass es darauf ankommt][indirekte Frage]]* und (iii) *[[es kommt auf][NP_{AKK}][an]]*.

⁸⁷ Problematisch erscheint hierbei die Tatsache, dass muttersprachliche Urteile bzgl. der (Nicht-) Akzeptabilität einer Konstruktion mitunter voneinander abweichen (vgl. Handwerker 2015: 149).

7. AUSBLICK

Im Rahmen der Arbeit wurden gebrauchsbasierte Sprachtheorien in den Fokus gerückt, die Äußerungen nicht mehr in regelmäßig und unregelmäßig einteilen, sondern diese einem Lexikon-Syntax-Kontinuum von mehr oder weniger abstrakt zuordnen. Spracherwerb erfolgt, indem Lernende mittels allgemein kognitiver Prozesse Gesetzmäßigkeiten aus Form-Bedeutungspaaren im Input ableiten und so ein Inventar von mehr oder weniger schematischen Konstruktionen aufbauen. Das Speichern partiell und vollständig formelhafter Sequenzen befähigt auf diese Weise sowohl Muttersprachler als auch Fremdsprachenlernende, natürliche Äußerungen schnell und flüssig zu produzieren. Zudem werden L2-Lernende auf unteren Niveaus dazu befähigt, kommunikative Handlungen auszuüben, wenngleich die in den Chunks enthaltenen grammatischen Informationen ihren Kenntnisstand überschreiten. Um diese Informationen für erwachsene Lernende nutzbar zu machen, müssen sie jedoch explizit vermittelt werden, da Lernende den Inhalt einer Äußerung verarbeiten, formseitige Besonderheiten jedoch nicht wahrnehmen.

Vom konkreten Sprachgebrauch ausgehend konnten in der Analyse eine Reihe von Chunks identifiziert werden, die von Lernenden auf fortgeschrittenen Niveaus verwendet werden. Dabei wurden sowohl feste als auch flexiblere Chunks mit Leerstellen analysiert. Anhand der unterschiedlichen, z.T. inkorrekt gebrauchten Füller, lassen sich akzeptable Füller, aber auch negative Evidenz und somit Restriktionen von Konstruktionen aufzeigen. Dies kann im Unterricht eingesetzt werden, um Lernende bei der kreativen Produktion von partiell schematischen Chunks zu unterstützen. Ferner wurden vergleichend für L1 und L2 Gebrauchspräferenzen bzw. *over-* und *underuse* bestimmter Strukturen der L2-Lernenden aufgedeckt, was entscheidend für einen Fremdsprachenunterricht ist, der sowohl die Form- als auch Funktionsseite von Chunks berücksichtigt.

Um die Aussagekraft der gewonnenen Erkenntnisse kritisch zu beleuchten, soll die angewandte Methodik nun hinsichtlich ihrer Reliabilität und Validität geprüft werden (vgl. Read/Nation 2004: 34f.). Während die externe Reliabilität durch klar beschriebene Verfahren der Datenerhebung und durch die Listen und Tabellen 1 bis 8, die im Anhang aufgeführt sind und alle extrahierten 3- und 4-Gramme enthalten, sodass die Identifikation der Chunks nachvollzogen werden kann, gegeben ist, könnte die interne Reliabilität optimiert werden. Idealerweise gäbe es mindestens zwei Sprachwissenschaftler, die unabhängig voneinander dieselben n-Gramme als Chunks identifizieren. Auch im Falle von (Nicht-)

Akzeptabilitätsurteilen in Bezug auf L2-Realisationen von schematischen Chunks sollten bestenfalls mehrere Muttersprachler befragt werden. Bemüht man das Kriterium „ganzheitlich gespeichert“ für Chunks, ist dies insofern problematisch, als es schwierig zu operationalisieren ist und somit eine geringe interne Validität aufweist. In der vorliegenden Arbeit wurden daher die Kriterien „kommunikative Einheit“, d.h. ohne zusätzliche, sprachliche Elemente verständlich, und ein Frequenzminimum von fünf angewandt, wobei der Faktor Intuition nicht ausgeschlossen werden konnte. Die externe Validität betrifft das Korpus, was in den meisten Fällen genügend aussagekräftige Beispiele enthielt, wobei ein größeres muttersprachliches Vergleichskorpus wünschenswert wäre. Zudem stellt sich die Frage, was als Kontrollkorpus dienen sollte – studentische Texte oder eventuell Expertentexte (vgl. Callies 2015: 40)? Auch Tippfehler bzw. bei handschriftlich verfassten Texten schwer leserliche Wörter, die in das Korpus übertragen wurden, können die gewonnenen Ergebnisse verfälscht haben.

Aus diesem Resümee ergeben sich folgende Desiderata für zukünftige Arbeiten: Bezüglich der Schwierigkeit, den Chunkstatus, bzw. wenn man von muttersprachlicher Kompetenz als eine strukturierte Ansammlung von Konstruktionen ausgeht, den Konstruktionsstatus einer Wortverbindung bei der Extraktion aus einem Korpus sicherzustellen, sollten zusätzlich Experimente durchgeführt werden, die eine ganzheitliche Speicherung bestätigen (Gilquin/Gries 2009). Allgemein werden objektivere und besser messbare Kriterien für die Auswahl kognitiv relevanter n-Gramme benötigt (vgl. Hein/Bubenhofer 2015: 203). Die Maße Dispersion und entropy können bei größeren Korpora angewandt werden, um frequente und regelmäßig vorkommende Chunks zu identifizieren und um die (un)ausgeglichene Verteilung von Füllern einer Leerstelle zu berechnen. Zur Identifizierung abstrakter Muster können anstelle von Lexemen frequente, gemeinsam auftretende Wortarten (POS-Tags) im Korpus gesucht werden. Das Verhältnis von Lexemen und grammatischen Konstruktionen kann dabei mittels *collostructional analyses* (Stefanowitsch/Gries 2003) untersucht werden. Auch die Suche nach diskontinuierlichen Chunks kann neue Erkenntnisse liefern.

Um Aussagen darüber treffen zu können, welche Chunks von Lernenden auf welchen Niveaus gebraucht werden, werden weitere Korpusanalysen benötigt, wobei die Korpora genauer auf die Niveaus der Lernenden abgestimmt sein müssen als dies beim Falko-Korpus der Fall ist. Daraus können zum einen neue Erkenntnisse über den Verlauf des Spracherwerbs gewonnen werden, zum anderen ergeben sich daraus Konsequenzen für den Unter-

richt und für das Testen von Deutsch als Fremdsprache. Ferner muss zwischen gesprochener und geschriebener Sprache unterschieden werden, denn diese werden unter anderen Bedingungen produziert.

Bezüglich der Bedeutungsseite von Chunks ist festzuhalten, dass sowohl Konstruktionsgrammatik als auch Phraseologie große Defizite im Bereich der Pragmatik beklagen, so dass abzuwarten ist, inwieweit zukünftige Studien in diesem Bereich neue Erkenntnisse für die DaF-Praxis bereithalten (vgl. Filatkina 2007: 152; vgl. Ziem 2015: 2).

Zuletzt müssen Lehrende für die zahlreichen Vorteile des Chunk-Lernens sensibilisiert werden, denn zumeist „[wird] [n]och immer [...] die Wiedergabe bzw. Paraphrase ‚mit eigenen Worten‘ höher bewertet als der Abruf von memorisierten [...] Konstruktionen“ (Aguado 2008: 57).

LITERATURVERZEICHNIS

- Abbot-Smith, K./Behrens, H. (2006): How Known Constructions Influence the Acquisition of Other Constructions: The German Passive and Future Constructions. In: *Cognitive Science* 30, 995–1026. DOI:10.1207/s15516709cog0000_61.
- Ágel, V. (2004): Phraseologismus als (valenz)syntaktischer Normalfall. In: K. Steyer (Hg.), *Wortverbindungen - Mehr Oder Weniger Fest*. Berlin/New York, 65–86.
- Aguado, K. (2008): Wie Beeinflussbar ist die Lersprachliche Entwicklung. In: *Fremdsprache Deutsch* 38, 53–58.
- (2013): Sind L2-Chunks Lebenslang Erlernbar oder Gibt es eine Sensible Phase für ihren Erfolgreichen Erwerb? In: A. Berndt (Hg.), *Fremdsprachen in der Perspektive Lebenslangen Lernens*. Frankfurt a. M./New York, 129–142.
- Ahrens, K. (1995): *The Mental Representation of Verbs*. University of California San Diego Dissertation.
- Ambridge, B./Rowland, C. F. (2009): Predicting Children's Errors with Negative Questions: Testing a Schema-Combination Account. In: *Cognitive Linguistics* 20/2, 225–266.
- Arnon, I./Snider, N. (2010): More than Words: Frequency Effects for Multi-Word Phrases. In: *Journal of Memory and Language* 62/1, 67–82.
- Baker, C.: FrameNet Project. Unter: <<https://framenet.icsi.berkeley.edu/fndrupal/>> (1.12.16).
- Bannard, C./Lieven, E. (2012): Formulaic Language in L1 Acquisition. In: *Annual Review of Applied Linguistics* 32, 3–16.
- Bannard, C./Matthews, D. (2008): Stored Word Sequences in Language Learning. The Effect of Familiarity on Children's Repetition of Four-Word Combinations. In: *Psychological Science* 19/3, 241–248.
- Barcroft, J./Wong, W. (2013): Input, Input Processing and Focus on Form. In: J. R. Herschensohn/M. Young-Scholten (Hg.), *The Cambridge Handbook of Second Language Acquisition*. Cambridge/New York, 627–647.
- Bardovi-Harlig, K. (1992): The Use of Adverbials and Natural Order in the Development of Temporal Expression. In: *International Review of Applied Linguistics in Language Teaching* 30/4, 299–320.
- Bates, E./Bretherton, I./Snyder, L. (1988): *From First Words to Grammar. Individual Differences and Dissociable Mechanisms*. Cambridge.
- Boas, H. Ch. (2003): *A Constructional Approach to Resultatives*. Stanford.
- (Hg.) (2010): *Contrastive Studies in Construction Grammar. Constructional Approaches to Language*. Amsterdam/Philadelphia.
- (2013): *Cognitive Construction Grammar*. In: Th. Hoffmann/G. Trousdale (Hg.), *The Oxford Handbook of Construction Grammar*. Oxford/New York, 233–252.
- Bransford, J. D./Barclay, J. R./Franks, J. J. (1972): Sentence Memory: A Constructive versus Interpretive Approach. In: *Cognitive Psychology* 3/2, 193–209.
- Burger, H. (2015): *Phraseologie: Eine Einführung am Beispiel des Deutschen*. Berlin.
- Bybee, J. (2008): Usage-Based Grammar and Second Language Acquisition. In: P. Robinson/N. C. Ellis (Hg.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York, 216–236.
- (2013): Usage-Based Theory and Exemplar Representations of Constructions. In: Th. Hoffmann/G. Trousdale (Hg.), *The Oxford Handbook of Construction Grammar*. Oxford/New York, 49–69.
- Bybee, J./Hopper, P. J. (Hg.) (2001): *Frequency and the Emergence of Linguistic Structure*. Amsterdam.
- Bybee, J./Thompson, S. (1997): Three Frequency Effects in Syntax. In: *The Annual Proceedings of the Berkeley Linguistics Society* 23, 378–388.
- Callies, M. (2015): Learner Corpus Methodology. In: S. Granger/G. Gilquin/F. Meunier (Hg.), *The Cambridge Handbook of Learner Corpus Research*. Cambridge, 35–55.
- Chan, A./Lieven, E./Tomasello, M. (2009): Children's Understanding of the Agent-Patient Relations in the Transitive Construction: Cross-Linguistic Comparisons between Cantonese, German, and English. In: *Cognitive Linguistics* 20/2, 267–300.
- Chomsky, N. (1959): A Review of B. F. Skinner's *Verbal Behavior*. In: *Language* 35/1, 26–58.
- (1965): *Aspects of the Theory of Syntax*. Cambridge, MA.

- (1981): Regeln und Repräsentationen. Frankfurt a. M.
- (1995): The Minimalist Program. Cambridge, MA.
- (2002): Syntactic Structures. Berlin/New York. Unter: <http://public.eblib.com/choice/publicfullrecord.aspx?p=773000>.
- Conklin, K./Schmitt, N. (2008): Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? In: *Applied Linguistics* 29/1, 72–89.
- Corder, S. P. (1967): The Significance of Learner's Errors. In: *International Review of Applied Linguistics in Language Teaching* 5/1, 161–170.
- Croft, W. (2001): *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford/New York.
- Croft, W./Cruse, D. A. (2004): *Cognitive Linguistics*. Cambridge/New York.
- Diehl, E./Christen, H./Leuenberger, S./Pelvat, I./Studer, Th. (Hg.) (2000): *Grammatikunterricht, Alles Für Der Katz? Untersuchungen zum Zweitspracherwerb Deutsch*. Tübingen.
- Diessel, H. (2013): Construction Grammar and First Language Acquisition. In: Th. Hoffmann/G. Trousdale (Hg.), *The Oxford Handbook of Construction Grammar*. Oxford/New York, 347–364.
- Dobrovolskij, D. (2011): Phraseologie und Konstruktionsgrammatik. In: A. Lasch/A. Ziem (Hg.), *Konstruktionsgrammatik. 3: Aktuelle Fragen und Lösungsansätze*. Tübingen, 111–130.
- Dumant, A./Granger, S.: Learner Corpora Around the World. <<https://www.uclouvain.be/en-cecl-lcworld.html>> (2.12.16).
- Durrant, P./Schmitt, N. (2009): To What Extent Do Native and Non-Native Writers Make Use of Collocations? In: *International Review of Applied Linguistics in Language Teaching* 47/2, 157–177.
- Elio, R./Anderson, J. R. (1984): The Effects of Information Order and Learning Mode on Schema Abstraction. In: *Memory & Cognition* 12/1, 20–30.
- Ellis, N. C. (2002): Frequency Effects in Language Processing. In: *Studies in Second Language Acquisition* 24/2, 143–188. DOI:10.1017/S0272263102002024.
- (2003): Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure. In: C. J. Doughty/M. H. Long (Hg.), *The Handbook of Second Language Acquisition*. Malden, MA, 63–103.
- (2007): The Associative-Cognitive CREED. In: B. VanPatten/J. Williams (Hg.), *Theories in Second Language Acquisition: An Introduction*. Mahwah, NJ, 77–95.
- (2008): Usage-Based and Form-Focused Language Acquisition. The Associative Learning of Constructions, Learned Attention, and the Limited L2 Endstate. In: P. Robinson/N. C. Ellis (Hg.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York, 372–405.
- (2009): Optimizing the Input: Frequency and Sampling in Usage-Based and Form-Focused Learning. In: M. H. Long/C. Doughty (Hg.), *The Handbook of Language Teaching*. Chichester/Malden, MA, 139–158.
- (2012): Frequency-based accounts of second language acquisition. In: S. M. Gass/A. Mackey (Hg.), *The Routledge handbook of second language acquisition*. London/New York, 193–210.
- (2013): Construction Grammar and Second Language Acquisition. In: Th. Hoffmann/G. Trousdale (Hg.), *The Oxford Handbook of Construction Grammar*. Oxford/New York, 365–378.
- Ellis, N. C./Ferreira-Junior, F. (2009): Constructions and Their Acquisition. Islands and the Distinctiveness of Their Occupancy. In: *Annual Review of Cognitive Linguistics* 7/1, 187–220.
- Ellis, N. C./Simpson-Vlach, R./Maynard, C. (2008): Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. In: *Tesol Quarterly* 42/3, 375–396.
- Ellis, N. C./Simpson-Vlach, R./Römer, U./O'Donnell, M. B./Wulff, S. (2015): Learner Corpora and Formulaic Language in Second Language Acquisition Research. In: S. Granger/G. Gilquin/F. Meunier (Hg.), *The Cambridge Handbook of Learner Corpus Research*. Cambridge, 357–378.
- Erman, B./Warren, B. (2000): The Idiom Principle and the Open Choice Principle. In: *Interdisciplinary Journal for the Study of Discourse* 20/1, 29–62.
- Falko (*ANNIS Corpus Search*). Unter: <<https://korpling.german.hu-berlin.de/falko-suche/>> (15.11.16).
- (*Metadaten der Teilnehmenden*). Unter: <<https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko>> (14.11.16).
- Fandrych, Ch./Tschirner, E. (2007): Korpuslinguistik und Deutsch als Fremdsprache. Ein Perspektivenwechsel. In: *Deutsch als Fremdsprache* 4, 195–204.

- Feilke, H. (2007): Syntaktische Aspekte der Phraseologie III: Construction Grammar und Verwandte Ansätze. In: H. Burger/D. Dobrovol'skij/P. Kühn/N. R. Norrick (Hg.), *Phraseologie. Ein Internationales Handbuch der Zeitgenössischen Forschung*, Halbbd. 1. Berlin/New York, 63–76.
- Filatkina, N. (2007): Pragmatische Beschreibungsansätze. In: H. Burger/D. Dobrovol'skij/P. Kühn/N. R. Norrick (Hg.), *Phraseologie. Ein Internationales Handbuch der Zeitgenössischen Forschung*, Halbbd. 1. Berlin/New York, 132–158.
- Filatkina, N./Steyer, K./Stumpf, S. (n.d.): *Formelhafte Sprache/Formulaic Language*. Berlin/New York.
- Fillmore, Ch. J. (1968): The Case for Case. In: E. Bach/R. T. Harms (Hg.), *Universals in Linguistic Theory*. New York et al., 1–88.
- Fillmore, Ch. J./Baker, C. (2015): A Frames Approach to Semantic Analysis. In: B. Heine/H. Narrog (Hg.), *The Oxford Handbook of Linguistic Analysis*. New York, 791–816.
- Fischer, K./Stefanowitsch, A. (2006): Konstruktionsgrammatik: Ein Überblick. In: K. Fischer/A. Stefanowitsch (Hg.), *Konstruktionsgrammatik. 1: Von der Anwendung zur Theorie*. Tübingen, 3–17.
- Forgács, T. (2009): Probleme der Klassifizierung phraseologischer Einheiten. In: C. Földes (Hg.), *Phraseologie Disziplinär und Interdisziplinär*. Tübingen, 131–143.
- Gentner, D./Medina, J. (1998): Similarity and the Development of Rules. In: *Cognition* 65/2, 263–297.
- Gilquin, G./Gries, S. Th. (2009): Corpora and Experimental Methods: A State-of-the-Art Review. In: *Corpus Linguistics and Linguistic Theory* 5/1, 1–26.
- Gläser, R. (1998): The Stylistic Potential of Phraseological Units in the Light of Genre Analysis. In: A. P. Cowie (Hg.), *Phraseology: Theory, Analysis, and Applications*. Oxford/New York, 125–143.
- Goldberg, A. E. (1995): *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago/London.
- (2006): *Constructions at Work: The Nature of Generalization in Language*. Oxford.
- (2013): Constructionist Approaches. In: Th. Hoffmann/G. Trousdale (Hg.), *The Oxford Handbook of Construction Grammar*. Oxford/New York, 15–31.
- Goldberg, A. E./Casenhiser, D. M. (2008): Construction Learning and Second Language Acquisition. In: P. Robinson/N. C. Ellis (Hg.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York, 197–215.
- Goldberg, A. E. /Casenhiser, D. M./Sethuraman, N. (2004): Learning Argument Structure Generalizations. In: *Cognitive Linguistics* 15/3, 289–316.
- Goldschneider, J. M./DeKeyser, R. M. (2001): Explaining the 'Natural Order of L2 Morpheme Acquisition' in English. A Meta-Analysis of Multiple Determinants. In: *Language Learning* 51/1, 1–50.
- Goschler, J. et al.: Arbeitskreis Konstruktionsgrammatik in Deutschland, Österreich und der Schweiz. Unter: <<https://konstruktionsgrammatik.wordpress.com>> (1.11.16).
- Granger, S. (1998): Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae. In: A. P. Cowie (Hg.), *Phraseology: Theory, Analysis, and Applications*. Oxford/New York, 145–160.
- (2002): A Bird's-Eye View of Learner Corpus Research. In: S. Granger/J. Hung/S. Petch-Tyson (Hg.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam/Philadelphia, 3–33.
- Granger, S./Paquot, M. (2008): Disentangling the Phraseological Web. In: S. Granger/F. Meunier (Hg.), *Phraseology: An Interdisciplinary Perspective*. Amsterdam/Philadelphia, 26–49.
- Gries, S. Th. (2008a): Phraseology and Linguistic Theory: A Brief Survey. In: S. Granger/F. Meunier (Hg.), *Phraseology: An Interdisciplinary Perspective*. Amsterdam/Philadelphia, 3–25.
- (2008b): Corpus-Based Methods in Analyses of Second Language Acquisition Data." In: P. Robinson/N. C. Ellis (Hg.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York, 406–431.
- (2008c): Dispersions and Adjusted Frequencies in Corpora. In: *International Journal of Corpus Linguistics* 13/4, 403–437.
- (2010): Dispersions and Adjusted Frequencies in Corpora: Further Explorations. In: S. Th. Gries/S. Wulff/M. Davies (Hg.), *Corpus Linguistic Applications: Current Studies, New Directions*. Amsterdam/New York, 197–212.
- Gries, S. Th./Ellis, N. C. (2015): Statistical Measures for Usage-Based Linguistics. In: *Language Learning* 65/1, 1–28. DOI: 10.1111/lang.12119.
- Gries, S. Th./Wulff, S. (2005): Do Foreign Language Learners Also Have Constructions? In: *Annual Review of Cognitive Linguistics* 3/1, 182–200.

- Grünhagen, T. (2015): Was Produzieren DaF-Lerner? - Zum Lernerkorpus FALKO und seinem Nutzen für die DaF-Vermittlung. In: B. Handwerker/R. Bäuerle/I. D. Reixa/B. Lübke (Hg.), *Zwischenräume: Lexikon und Grammatik im Deutschen als Fremdsprache*. Baltmannsweiler, 97–114.
- Häcki Buhofer, A./Dräger, M./Meier, S./Roth, T. (Hg.) (2014): *Feste Wortverbindungen des Deutschen: Kollokationenwörterbuch für den Alltag*. Tübingen.
- Handwerker, B. (2008): Chunks und Konstruktionen — Zur Integration von Lerntheoretischem und Grammatischem Ansatz. In: *Estudios Filológicos Alemanes* 15, 49–64.
- (2010): Phraseologismen und Kollokationen. In: Ch. Fandrych et al. (Hg.), *Deutsch Als Fremd- und Zweitsprache. Ein Internationales Handbuch*. Berlin/New York, 246–255.
- (2015): Konstruktionen im L2-Lernformat. Orts- und Zustandsveränderungen in der Rezeption und Produktion des Deutschen als Fremdsprache. In: S. Engelberg/M. Meliss/K. Proost/E. Winkler (Hg.), *Argumentstruktur Zwischen Valenz und Konstruktion*. Tübingen, 127–153.
- Heinemann, M./Heinemann, W. (2002): *Grundlagen der Textlinguistik: Interaktion - Text - Diskurs*. Tübingen.
- Hein, K./Bubenhof, N. (2015): Korpuslinguistik Konstruktionsgrammatisch. In: A. Ziem/A. Lasch (Hg.), *Konstruktionen als Soziale Konventionen und Kognitive Routinen*. Tübingen, 179–206.
- Helbig, G./Buscha, J. (2001): *Deutsche Grammatik: Ein Handbuch für den Ausländerunterricht*. Berlin.
- Hoey, M. (2004): The Textual Priming of Lexis. In: G. Aston/S. Bernardini/D. Stewart (Hg.), *Corpora and Language Learners*. Amsterdam/Philadelphia, 21–41.
- Holme, R. (2010): A Construction Grammar for the Classroom. In: *International Review of Applied Linguistics in Language Teaching* 48/4, 355–377.
- Jiang, N./Nekrasova, T. M. (2007): The Processing of Formulaic Sequences by Second Language Speakers. In: *The Modern Language Journal* 91/3, 433–445.
- Kay, P. (2005): Argument Structure Constructions and the Argument-Adjunct Distinction. In: M. Fried/H. C. Boas (Hg.), *Grammatical Constructions: Back to the Roots*. Amsterdam/Philadelphia, 71–98.
- Kay, P./Fillmore, Ch. J. (1999): Grammatical Constructions and Linguistic Generalizations: The What's X Doing Y? Construction. In: *Language* 1/75, 1–33.
- Lakoff, G. (1990): The Invariance Hypothesis: Is Abstract Reason Based on Image-Schemas? In: *Cognitive Linguistics* 1/1, 39–74. DOI:10.1515/cogl.1990.1.1.39.
- Langacker, R. W. (2013): *Essentials of Cognitive Grammar*. Oxford/New York.
- Lewis, M. (2008): *Implementing the Lexical Approach. Putting Theory into Practice*. Boston.
- Lieven, E./Tomasello, M. (2008): Children's First Language Acquisition from a Usage-Based Perspective. In: P. Robinson/N. C. Ellis (Hg.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York, 168–196.
- Lieven, E./Salomo, D./Tomasello, M. (2009): Two-Year-Old Children's Production of Multiword Utterances: A Usage-Based Analysis. In: *Cognitive Linguistics* 20/3, 481–507.
- Lüdeling, A./Dookittle, S./Hirschmann, H./Schmidt, K./Walter, M. (2008): Das Lernerkorpus Falko. In: *Deutsch Als Fremdsprache* 45/2, 67–73.
- Lüdeling, A./Reznicek, M./Krummes, C./Schwantuschke, F./Walter, M./Schmidt, K./Hirschmann, H./Andreas, T. (2012): *Das Falko-Handbuch. Korpusaufbau und Annotationen. Version 2.01*. Unter: https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/FalkoHandbuchV2/at_download/file.
- MacWhinney, B./Pléh, C./Bates, E. (1985): The Development of Sentence Interpretation in Hungarian. In: *Cognitive Psychology* 17/2, 178–209.
- Marschner, C./Langer, S. (2006): *Statistische Methoden in der Sprachverarbeitung*. Unter: http://www.cis.uni-muenchen.de/~micha/kurse/statistikSS2007/literatur/Statistik_Skript_Langer_Marschner.pdf.
- Matthews, D./Bannard, C. (2010): Children's Production of Unfamiliar Word Sequences Is Predicted by Positional Variability and Latent Classes in a Large Sample of Child-Directed Speech. In: *Cognitive Science* 34/3, 465–488.
- McEnery, T./Wilson, A. (2001): *Corpus Linguistics: An Introduction*. Edinburgh.
- Miller, G. A. (1956): The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. In: *Psychological Review* 63/2, 81–97.

- Moder, C. L. (1992): Productivity and Categorization in Morphological Classes. Dissertation.
- Myles, F. (2004): From Data to Theory: The Over-Representation of Linguistic Knowledge in SLA. In: Transactions of the Philological Society 102/2, 139–168.
- Myles, F./Mitchell, R./Hooper, J. (1999): Interrogative Chunks in French L2. A Basis for Creative Construction? In: Studies in Second Language Acquisition 21/1, 49–80.
- Nattinger, J. R./DeCarrico, J. S. (1992): Lexical Phrases and Language Teaching. Oxford.
- Nosofsky, R. M. (1988): Similarity, Frequency, and Category Representations. In: Journal of Experimental Psychology: Learning, Memory, and Cognition 14/1, 54–65.
- Odlin, T. (2003): Cross-Linguistic Influence. In: C. J. Doughty/M. H. Long (Hg.), The Handbook of Second Language Acquisition. Oxford, 436–486.
- O'Donnell, M. B./Römer, U./Ellis, N. C. (2013): The Development of Formulaic Sequences in First and Second Language Writing: Investigating Effects of Frequency, Association, and Native Norm. In: International Journal of Corpus Linguistics 18/1, 83–108.
- Pawley, A. (1986): On Speech Formulas and Linguistic Competence. In: Kansas Working Papers in Linguistics 11, 57–87. Unter: <https://kuscholarworks.ku.edu/bitstream/handle/1808/566/ling.wp.v11.n1.paper5.pdf?sequence=1&isAllowed=y>.
- Pawley, A./Syder, F. H. (1983): Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency. In: J. C. Richards/R. Schmidt (Hg.), Language and Communication. London, 191–226.
- Pica, T. (1983): Adult Acquisition of English as a Second Language under Different Conditions of Exposure. In: Language Learning 33/4, 465–497.
- Read, J./Nation, P. (2004): Measurement of Formulaic Sequences. In: N. Schmitt (Hg.), Formulaic Sequences: Acquisition, Processing, and Use. Amsterdam/Philadelphia, 23–35.
- Robinson, P. J./Ellis, N. C. (Hg.) (2008): Handbook of Cognitive Linguistics and Second Language Acquisition. New York.
- (2008): An Introduction to Cognitive Linguistics, Second Language Acquisition, and Language Instruction. In: P. Robinson/N. C. Ellis (Hg.), Handbook of Cognitive Linguistics and Second Language Acquisition. New York.
- Rostila, J. (2012): Konstruktionsgrammatik: Innovative Wege für den DaF-Unterricht, Insbesondere den Grammatikunterricht. In: German as a Foreign Language 2–3, 216–237.
- Saussure, F. de (1916): Grundfragen der Allgemeinen Sprachwissenschaft. In: L. Hoffmann (Hg.), Sprachwissenschaft. Ein Reader (2000). Berlin, 32–50.
- Schmitt, N./Carter, R. (2004): Formulaic Sequences in Action: An Introduction. In: N. Schmitt (Hg.), Formulaic Sequences: Acquisition, Processing, and Use. Amsterdam/Philadelphia, 1–22.
- Schwarz, M. (2008): Einführung in die Kognitive Linguistik. Tübingen.
- Sharwood Smith, M. (1993): Input Enhancement in Instructed SLA. In: Studies in Second Language Acquisition 15/2, 165–179.
- Simpson-Vlach, R./Ellis, N. C. (2009): Formulaic Language in Native Speakers: Triangulating Psycholinguistics, Corpus Linguistics, and Education. In: Corpus Linguistics and Linguistic Theory 5/1, 61–78. DOI:<http://dx.doi.org/10.1515/CLLT.2009.003>.
- Sinclair, J. (1987): Looking up: An Account of the COBUILD Project in Lexical Computing. London.
- (1991): Corpus, Concordance, Collocation. Oxford.
- Stefanowitsch, A. (2011a): Keine Grammatik Ohne Konstruktion: Ein Logisch-ökonomisches Argument für die Konstruktionsgrammatik. In: S. Engelberg/A. Holler/K. Proost (Hg.), Sprachliches Wissen zwischen Lexikon und Grammatik. Berlin/Boston, 181–210.
- (2011b): Konstruktionsgrammatik und Grammatiktheorie. In: A. Lasch/A. Ziem (Hg.), Konstruktionsgrammatik. 3: Aktuelle Fragen und Lösungsansätze. Tübingen, 11–25.
- Stefanowitsch, A./Gries, S. Th. (2003): Collocations: Investigating the Interaction of Words and Constructions. In: International Journal of Corpus Linguistics 8/2, 209–243.
- Stein, S. (1994): Neuere Literatur zur Phraseologie und zu Ritualisierter Sprache. In: Deutsche Sprache 22/2, 152–180.
- (1995): Formelhafte Sprache: Untersuchungen zu Ihren Pragmatischen und Kognitiven Funktionen im Gegenwärtigen Deutsch. Frankfurt a. M./New York.

- (2007): Mündlichkeit und Schriftlichkeit aus Phraseologischer Perspektive. In: H. Burger/D. Dobrovol'skij/P. Kühn/N. R. Norrick (Hg.), *Phraseologie. Ein Internationales Handbuch der Zeitgenössischen Forschung*, Halbbd. 1. Berlin/New York, 220–236.
- Steyer, K. (2003): Korpus, Statistik, Kookkurrenz. Lässt Sich Idiomatisches 'Berechnen'? In: H. Burger/A. Häcki Buhofer/G. Greciano (Hg.), *Flut von Texten - Vielfalt der Kulturen. Ascona 2001 Zur Methodologie und Kulturspezifität der Phraseologie*. Baltmannsweiler, 33–46.
- (2013): *Usuelle Wortverbindungen: Zentrale Muster des Sprachgebrauchs aus Korpusanalytischer Sicht*. Tübingen.
- Stoll, S./Abbot-Smith/K./Lieven/E. (2009): Lexically Restricted Utterances in Russian, German, and English Child-Directed Speech. In: *Cognitive Science* 33/1, 75–103.
- Stubbs, M. (2007): Quantitative Data on Multi-Word Sequences in English: The Case of the Word World. In: M. Hoey/M. Mahlberg/M. Stubbs/ W. Teubert (Hg.), *Text, Discourse and Corpora: Theory and Analysis*. London/New York, 163–189.
- Stumpf, S. (2015): *Formelhafte (Ir-)regularitäten: Korpuslinguistische Befunde und Sprachtheoretische Überlegungen*. Frankfurt a. M. /New York.
- Tomasello, M. (2003): *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge.
- (2006): Konstruktionsgrammatik und früher Erstspracherwerb. In: K. Fischer/A. Stefanowitsch (Hg.), *Konstruktionsgrammatik. 1: Von der Anwendung zur Theorie*. Tübingen, 19–37.
- Tremblay, A./Derwing, B./Libben, G./Westbury, C. (2011): Processing Advantages of Lexical Bundles: Evidence From Self-Paced Reading and Sentence Recall Tasks. In: *Language Learning* 61/2, 569–613. DOI:10.1111/j.1467-9922.2010.00622.x.
- Tschirner, E. (2000): Digitale Klassenzimmer: Sieben Thesen zum Erwerb mündlicher Handlungsfähigkeit. In: E. Tschirner/H. Funk/M. König (Hg.), *Schnittstellen: Lehrwerke zwischen alten und neuen Medien*. Berlin, 66–93.
- (2005): Korpora, Häufigkeitslisten, Wortschatzerwerb. In: E. Tschirner/A. Heine/M. Hennig (Hg.), *Deutsch Als Fremdsprache-Konturen und Perspektiven eines Fachs*. München, 133–149.
- VanPatten, B. (2004): Input Processing in SLA. In: B. VanPatten (Hg.), *Processing Instruction: Theory, Research, and Commentary*. Mahwah, NJ, 5–31.
- Wegener, H. (2010): Morphologie: Flexion. In: Ch. Fandrych et al. (Hg.), *Deutsch Als Fremd- und Zweitsprache. Ein Internationales Handbuch*. Berlin/New York, 206–216.
- Welke, K. (2013): Konstruktionsgrammatik (KxG) und Deutsch als Fremdsprache (DaF). In: *Deutsch Als Fremdsprache* 50/1, 19–27.
- White, L. (2007): Linguistic Theory, Universal Grammar, and Second Language Acquisition. In: B. VanPatten/J. Williams (Hg.), *Theories in Second Language Acquisition: An Introduction*. Mahwah, NJ, 37–55.
- Wittek, A./Tomasello, M. (2005): German-Speaking Children's Productivity with Syntactic Constructions and Case Morphology: Local Cues Act Locally. In: *First Language* 25/1, 103–125.
- Wotjak, B./Heine, A. (2007): Syntaktische Aspekte der Phraseologie I: Valenztheoretische Ansätze. In: H. Burger/D. Dobrovol'skij/P. Kühn/N. R. Norrick (Hg.), *Phraseologie. Ein Internationales Handbuch der Zeitgenössischen Forschung*, Halbbd. 1. Berlin/New York, 41–53.
- Wray, A. (2002): *Formulaic Language and the Lexicon*. Cambridge.
- Wray, A./Perkins, M. R. (2000): The Functions of Formulaic Language: An Integrated Model. In: *Language & Communication* 20/1, 1–28.
- Yorio, C. A. (1989): Idiomaticity as an Indicator of Second Language Proficiency. In: K. Hyltenstam/L. K. Obler (Hg.), *Bilingualism across the Lifespan: Aspects of Acquisition, Maturity and Loss*. Cambridge, 55–72.
- Ziem, A. (2014): Konstruktionsgrammatische Konzepte eines Konstruktions. In: A. Ziem/A. Lasch (Hg.), *Grammatik als Netzwerk von Konstruktionen. Sprachliches Wissen im Fokus der Konstruktionsgrammatik*. Berlin/New York, 15–34.
- (2015): Desiderata und Perspektiven einer Social Construction Grammar. In: A. Ziem/A. Lasch (Hg.), *Konstruktionsgrammatik IV. Konstruktionen als Soziale Konventionen und Kognitive Routinen*. Tübingen, 1–21.
- Ziem, A./Lasch, A. (2013): *Konstruktionsgrammatik: Konzepte und Grundlagen Gebrauchsbasierter Ansätze*. Berlin/Boston.

Zipf, G. K. (1935): The Psycho-Biology of Language. Boston.

Liste 1. L2_n3_mostcommon

[('die wirkliche welt', 299),
 ('interessen der frauen', 242),
 ('auf die wirkliche', 238),
 ('den interessen der', 214),
 ('von geringem wert', 199),
 ('f\xfcr die gesellschaft', 194),
 ('nicht auf die', 162),
 ('die studenten nicht', 157),
 ('wirkliche welt vor', 152),
 ('der frauen mehr', 146),
 ('studenten nicht auf', 145),
 ('die meisten
 universit\tsabschl\xfcse', 143),
 ('bereiten die studenten', 133),
 ('meiner meinung nach', 132),
 ('sind nicht praxisorientiert', 131),
 ('und bereiten die', 129),
 ('der feminismus hat', 129),
 ('nicht praxisorientiert und', 125),
 ('frauen mehr geschadet', 125),
 ('sind deswegen von', 123),
 ('sie sind deswegen', 123),
 ('mehr geschadet als', 122),
 ('praxisorientiert und bereiten',
 118),
 ('geschadet als gen\xtzt', 116),
 ('in der gesellschaft', 115),
 ('universit\tsabschl\xfcse sind
 nicht', 112),
 ('deswegen von geringem', 112),
 ('sich nicht aus', 112),
 ('meisten
 universit\tsabschl\xfcse sind',
 108),
 ('kriminalit\xt zahlt sich', 108),
 ('zahlt sich nicht', 104),
 ('feminismus hat den', 100),
 ('hat den interessen', 99),
 ('die finanzielle entlohnung', 99),
 ('an der universit\xt', 96),
 ('dass der feminismus', 83),
 ('entlohnung eines menschen', 83),
 ('finanzielle entlohnung eines', 83),

('die gesellschaft geleistet', 76),
 ('gesellschaft geleistet hat', 72),
 ('sie f\xfcr die', 68),
 ('der wirklichen welt', 67),
 ('dem beitrags entsprechen', 67),
 ('feminismus den interessen', 66),
 ('entsprechen den er', 63),
 ('beitrags entsprechen den', 62),
 ('der meinung dass', 60),
 ('den er sie', 60),
 ('der feminismus den', 60),
 ('er sie f\xfcr', 57),
 ('kriminalit\xt sich nicht', 57),
 ('an der uni', 54),
 ('der anderen seite', 51),
 ('eines menschen sollte', 51),
 ('nicht praxisorientiert sind', 50),
 ('sollte dem beitrags', 50),
 ('es gibt viele', 49),
 ('menschen sollte dem', 47),
 ('auf der anderen', 47),
 ('dass die studenten', 46),
 ('in der welt', 45),
 ('ich glaube dass', 45),
 ('in der wirklichen', 43),
 ('m\enne und frauen', 43),
 ('sagen dass die', 42),
 ('dass kriminalit\xt sich', 42),
 ('es gibt auch', 41),
 ('sich nicht auszahlt', 40),
 ('dass die frauen', 39),
 ('dass die meisten', 39),
 ('zu sagen dass', 37),
 ('frauen in der', 37),
 ('man sagen dass', 36),
 ('ist es nicht', 35),
 ('auf jeden fall', 35),
 ('und es ist', 35),
 ('frauen und m\enne', 34),
 ('aber es gibt', 34),
 ('die leute die', 34),
 ('in der arbeitswelt', 33),
 ('in der vergangenheit', 32),
 ('man in der', 32),
 ('geld zu verdienen', 32),

('studenten auf die', 31),
 ('als gen\xtzt hat', 31),
 ('aber es ist', 31),
 ('die kriminalit\xt sich', 31),
 ('und es gibt', 31),
 ('die studenten die', 31),
 ('sagen dass es', 31),
 ('hat der feminismus', 31),
 ('die frauen die', 30),
 ('in der schule', 30),
 ('und so weiter', 30),
 ('in bezug auf', 29),
 ('die interessen der', 29),
 ('es ist auch', 29),
 ('die studenten auf', 29),
 ('gibt es viele', 29),
 ('universit\tsabschl\xfcse nicht
 praxisorientiert', 29),
 ('es ist nicht', 28),
 ('wie die m\enne', 28),
 ('zu hause bleiben', 28),
 ('f\xfcr die frauen', 28),
 ('glaube ich dass', 28),
 ('in der heutigen', 28),
 ('in der zukunft', 27),
 ('haben die frauen', 27),
 ('wenn man sich', 27),
 ('nicht nur die', 27),
 ('nach dem studium', 27),
 ('nicht so viel', 26),
 ('wenn man eine', 26),
 ('wirkliche welt vorbereitet', 26),
 ('wenn man ein', 26),
 ('geringem wert sind', 26),
 ('einen job zu', 25),
 ('meinung nach ist', 25),
 ('wie zum beispiel', 25),
 ('in unserer gesellschaft', 25),
 ('und wenn man', 25),
 ('der feminismus ist', 24),
 ('in diesem fall', 24),
 ('in der universit\xt', 23),
 ('wenn man die', 23),
 ('der frauen geschadet', 23),
 ('bei der arbeit', 23),

(u'dass es nicht', 22),
(u'wirkliche welt vorbereiten', 22),
(u'aber wenn man', 22),
(u'dass frauen die', 22),
(u'ich sagen dass', 22),
(u'dass die kriminalit\xe4t', 22),
(u'f\xfc die studenten', 22),
(u'in den meisten', 22),
(u'was sie wollen', 21),
(u'ist dass die', 21),
(u'kriminalit\xe4t sich auszahlt', 21),
(u'kann man nicht', 21),
(u'sagen dass der', 21),
(u'in den letzten', 21),
(u'die mehrheit der', 21),
(u'f\xfc die wirkliche', 21),
(u'in den usa', 21),
(u'im vergleich zu', 21),
(u'nicht in der', 20),
(u'in der praxis', 20),
(u'dass die universit\xe4tsabschl\xfcsse', 20),
(u'gibt es auch', 20),
(u'dass wenn man', 20),
(u'man zum beispiel', 20),
(u'zu machen und', 20),
(u'nicht von geringem', 20),
(u'viele leute die', 20),
(u'wenn die studenten', 20),
(u'die frauen in', 20),
(u'rolle in der', 20),
(u'wichtig f\xfc die', 19),
(u'von der gesellschaft', 19),
(u'in diesem aufsatz', 19),
(u'es ist ein', 19),
(u'am ende des', 19),
(u'die frauen nicht', 19),
(u'und sie haben', 19),
(u'ich bin der', 19),
(u'praxisorientiert sind und', 19),
(u'nicht der fall', 19),
(u'sie in der', 19),
(u'ist es aber', 19),
(u'arbeit zu finden', 19),

(u'beitrag f\xfc die', 19),
(u'job zu bekommen', 19),
(u'zum beispiel wenn', 19),
(u'f\xfc die zukunft', 19),
(u'ein großes problem', 18),
(u'der frauen in', 18),
(u'gibt es ein', 18),
(u'bin der meinung', 18),
(u'gibt es eine', 18),
(u'ich denke dass', 18),
(u'kann man sagen', 18),
(u'die frage ist', 18),
(u'praktikum zu machen', 18),
(u'zum beispiel in', 18),
(u'die meisten studenten', 18),
(u'es gibt aber', 18),
(u'während des studiums', 18),
(u'es so viele', 18),
(u'es ist aber', 17),
(u'job zu finden', 17),
(u'ist nicht nur', 17),
(u'dass man nicht', 17),
(u'interesse der frauen', 17),
(u'im bezug auf', 17),
(u'zu tun hat', 17),
(u'gibt es immer', 17),
(u'es ist eine', 17),
(u'wenn man etwas', 17),
(u'die gesellschaft zu', 17),
(u'gesagt werden dass', 17),
(u'universit\xe4tsabschl\xfcsse von geringem', 17),
(u'k\xfcnte man sagen', 17),
(u'die in der', 16),
(u'wenn man einen', 16),
(u'der frauen gen\xfcgt', 16),
(u'der universit\xe4t zu', 16),
(u'meinung dass die', 16),
(u'dass die finanzielle', 16),
(u'k\xfcnte sagen dass', 16),
(u'meine meinung nach', 16),
(u'eines menschen dem', 16),
(u'ist dass man', 16),
(u'menschen dem beitrag', 16),
(u'man k\xfcnte sagen', 16),

(u'und dass sie', 16),
(u'deshalb ist es', 16),
(u'zum beispiel die', 16),
(u'gibt es noch', 16),
(u'sondern auch die', 16),
(u'um die kinder', 16),
(u'wenn sie nicht', 16),
(u'in die wirkliche', 15),
(u'man an der', 15),
(u'nicht nur f\xfc', 15),
(u'die mehrheit von', 15),
(u'mehr gen\xfcgt als', 15),
(u'wie kann man', 15),
(u'die rolle der', 15),
(u'nicht nur um', 15),
(u'nicht so gut', 15),
(u'zum beispiel ein', 15),
(u'im vergleich mit', 15),
(u'ist es auch', 15),
(u'frauen mehr gen\xfcgt', 15),
(u'der gesellschaft zu', 15),
(u'meinung nach sind', 15),
(u'und sie k\xfcnnen', 15),
(u'f\xfc sich selbst', 15),
(u'das leben der', 15),
(u'um geld zu', 15),
(u'die gleiche rechte', 14),
(u'sagen dass kriminalit\xe4t', 14),
(u'der heutigen gesellschaft', 14),
(u'ist es klar', 14),
(u'es gibt noch', 14),
(u'gibt es die', 14),
(u'und wenn sie', 14),
(u'frauen geschadet hat', 14),
(u'den letzten jahren', 14),
(u'und die studenten', 14),
(u'die meisten universit\xe4tsabschl\xfcsse', 14),
(u'und man muss', 14),
(u'und das ist', 14),
(u'ist klar dass', 14),
(u'und sie sind', 14),
(u'ein praktikum zu', 14),
(u'wenn eine frau', 14),
(u'muss man sich', 14),

(u'es ist sehr', 14),
 (u'dass sie die', 14),
 (u'meisten
 universit\xe4tsabschl\xfcsse nicht',
 14),
 (u'es ist klar', 14),
 (u'man kann nicht', 14),
 (u'auch wenn man', 14),
 (u'die frage ob', 14),
 (u'dass sie nicht', 14),
 (u'behaupten dass die', 14),
 (u'm\enne und frauen', 14),
 (u'auf der stra\se', 14),
 (u'es nicht so', 14),
 (u'man muss nicht', 14),
 (u'an der wirtschaftsuniversität',
 14),
 (u'nicht aus weil', 14),
 (u'das ist eine', 14),
 (u'nicht vergessen dass', 14),
 (u'ist es möglich', 14),
 (u'und nicht nur', 14),
 (u'wenn man in', 14),
 (u'sind der meinung', 14),
 (u'wegen des feminismus', 13),
 (u'nicht so einfach', 13),
 (u'zu bekommen und', 13),
 (u'sind und dass', 13),
 (u'frauen die gleiche', 13),
 (u'und die frauen', 13),
 (u'nach dem abschluss', 13),
 (u'für viele leute', 13),
 (u'und in der', 13),
 (u'zu finden und', 13),
 (u'es gibt ein', 13),
 (u'gen\stzt als geschadet', 13),
 (u'beispiel wenn man', 13),
 (u'der feminismus in', 13),
 (u'so viel geld', 13),
 (u'kann man sich', 13),
 (u'ich der meinung', 13),
 (u'das problem ist', 13),
 (u'die neuen studiengänge', 13),
 (u'sagen dass sie', 13),
 (u'der universit\st studieren', 13),

(u'viel mehr als', 13),
 (u'von geringen wert', 13),
 (u'muss man auch', 13),
 (u'vorbereitung auf die', 13),
 (u'auch in der', 13),
 (u'die interesse der', 13),
 (u'der einen seite', 13),
 (u'wenn man nicht', 13),
 (u'nicht für die', 13),
 (u'ist und dass', 13),
 (u'dass kriminalit\st zahlt', 13),
 (u'2 die meisten', 12),
 (u'für die', 12),
 (u'die tatsache dass', 12),
 (u'in der arbeit', 12),
 (u'es kann sein', 12),
 (u'für die arbeitswelt', 12),
 (u'es klar dass', 12),
 (u'einen beruf zu', 12),
 (u'ist es ganz', 12),
 (u'ab und zu', 12),
 (u'nicht sagen dass', 12),
 (u'die gesellschaft ist', 12),
 (u'vor allem die', 12),
 (u'an einer universität', 12),
 (u'sie haben keine', 12),
 (u'geld für die', 12),
 (u'heutzutage gibt es', 12),
 (u'das wort feminismus', 12),
 (u'zwischen m\enne und', 12),
 (u'denken dass es', 12),
 (u'es gibt keine', 12),
 (u'die universit\st\xfcsse
 nicht', 12),
 (u'ob der feminismus', 12),
 (u'eine gute arbeit', 12),
 (u'aber man muss', 12),
 (u'rechte der frauen', 12),
 (u'das neue system', 12),
 (u'in der uni', 12),
 (u'es sich nicht', 12),
 (u'geringem wert ist', 12),
 (u'es gibt immer', 12),
 (u'ist die kriminalit\st', 12),
 (u'im laufe der', 12),

(u'man in einer', 12),
 (u'man muss auch', 12),
 (u'ist sehr wichtig', 12),
 (u'es immer noch', 12),
 (u'ist es wichtig', 12),
 (u'aber das ist', 11)

Liste 2. L2_n4_mostcommon

[(u'auf die wirkliche welt', 235),
(u'den interessen der frauen', 210),
(u'nicht auf die wirkliche', 150),
(u'die wirkliche welt vor', 149),
(u'die studenten nicht auf', 142),
(u'studenten nicht auf die', 140),
(u'interessen der frauen mehr', 138),
(u'bereiten die studenten nicht',
125),
(u'und bereiten die studenten', 125),
(u'sie sind deswegen von', 122),
(u'sind nicht praxisorientiert und',
120),
(u'der frauen mehr geschadet', 120),
(u'praxisorientiert und bereiten die',
117),
(u'frauen mehr geschadet als', 117),
(u'nicht praxisorientiert und
bereiten', 114),
(u'mehr geschadet als gen\xftzt',
113),
(u'deswegen von geringem wert', 111),
(u'sind deswegen von geringem', 110),
(u'meisten
universit\xftsabschl\xfcsse sind
nicht', 106),
(u'universit\xftsabschl\xfcsse sind
nicht praxisorientiert', 106),
(u'die meisten
universit\xftsabschl\xfcsse sind',
105),
(u'zahlt sich nicht aus', 98),
(u'hat den interessen der', 97),
(u'kriminalit\xft sich nicht',
96),
(u'feminismus hat den interessen',
94),
(u'der feminismus hat den', 94),
(u'die finanzielle entlohnung eines',
81),
(u'finanzielle entlohnung eines
menschen', 75),
(u'f\xfr die gesellschaft geleistet',

73),
(u'die gesellschaft geleistet hat',
71),
(u'beitrag entsprechen den er', 62),
(u'sie f\xfr die gesellschaft', 62),
(u'feminismus den interessen der',
61),
(u'dem beitrag entsprechen den', 59),
(u'den er sie f\xfr', 57),
(u'er sie f\xfr die', 56),
(u'der feminismus den interessen',
52),
(u'entsprechen den er sie', 52),
(u'entlohnung eines menschen sollte',
51),
(u'sollte dem beitrag entsprechen',
49),
(u'eines menschen sollte dem', 47),
(u'auf der anderen seite', 47),
(u'menschen sollte dem beitrag', 47),
(u'in der wirklichen welt', 42),
(u'kriminalit\xft sich nicht
auszahlt', 36),
(u'dass der feminismus den', 35),
(u'dass kriminalit\xft sich nicht',
30),
(u'geschadet als gen\xftzt hat', 27),
(u'von geringem wert sind', 25),
(u'die interessen der frauen', 24),
(u'die wirkliche welt vorbereitet',
24),
(u'universit\xftsabschl\xfcsse nicht
praxisorientiert sind', 23),
(u'die wirkliche welt vorbereiten',
22),
(u'studenten auf die wirkliche', 22),
(u'die studenten auf die', 22),
(u'interessen der frauen geschadet',
21),
(u'meiner meinung nach ist', 21),
(u'nicht von geringem wert', 20),
(u'f\xfr die wirkliche welt', 20),
(u'dass die meisten
universit\xftsabschl\xfcsse', 19),
(u'beitrag f\xfr die gesellschaft',

18),
(u'bin der meinung dass', 17),
(u'ich bin der meinung', 17),
(u'universit\xftsabschl\xfcsse von
geringem wert', 17),
(u'entlohnung eines menschen dem',
16),
(u'dass die finanzielle entlohnung',
16),
(u'eines menschen dem beitrag', 15),
(u'der meinung dass die', 15),
(u'dass die kriminalit\xft sich',
15),
(u'nicht praxisorientiert sind und',
15),
(u'sagen dass der feminismus', 15),
(u'frauen mehr gen\xftzt als', 14),
(u'kann man sagen dass', 14),
(u'man könnte sagen dass', 14),
(u'in den letzten jahren', 14),
(u'k\xftnte man sagen dass', 14),
(u'die meisten
universit\xftsabschl\xfcsse nicht',
13),
(u'in der heutigen gesellschaft', 13),
(u'mehr gen\xftzt als geschadet',
13),
(u'der frauen mehr gen\xftzt', 13),
(u'zum beispiel wenn man', 13),
(u'es ist klar dass', 13),
(u'interessen der frauen gen\xftzt',
13),
(u'sind der meinung dass', 13),
(u'in die wirkliche welt', 13),
(u'er f\xfr die gesellschaft', 12),
(u'rolle in der gesellschaft', 12),
(u'ein praktikum zu machen', 12),
(u'die kriminalit\xft sich auszahlt',
12),
(u'meiner meinung nach sind', 12),
(u'der frauen in der', 12),
(u'zwischen m\xennern und frauen',
12),
(u'man an der universit\xft', 11),
(u'meisten

universit\xe4tsabschl\xfcsse nicht
 praxisorientiert', 11),
 (u'in der gesellschaft zu', 11),
 (u'menschen dem beitrag entsprechen',
 11),
 (u'hat der feminismus den', 11),
 (u'bin ich der meinung', 11),
 (u'auf der einen seite', 11),
 (u'an der universität studieren', 11),
 (u'von geringem wert ist', 11),
 (u'ich der meinung dass', 11),
 (u'zwischen m\ex4nner und frauen',
 11),
 (u'f\xfcfr die gesellschaft leisten',
 11),
 (u'kriminalit\ex4t sich nicht aus',
 10),
 (u'ich glaube dass die', 10),
 (u'es kann sein dass', 10),
 (u'der frauen geschadet hat', 10),
 (u'die kriminalit\ex4t sich nicht',
 10),
 (u'2 die meisten
 universit\ex4tsabschl\xfcsse', 10),
 (u'dass frauen und m\ex4nner', 10),
 (u'vorbereitung auf die wirkliche',
 10),
 (u'zu hause bleiben und', 10),
 (u'die interesse der frauen', 10),
 (u'den er f\xfcfr die', 10),
 (u'f\xfcfr die gesellschaft ist', 10),
 (u'frauen in der gesellschaft', 10),
 (u'die meisten
 universit\ex4tsabschl\xfcsse sind', 10),
 (u'f\xfcfr die gesellschaft machen',
 9),
 (u'das leben der frauen', 9),
 (u'ist es klar dass', 9),
 (u'meinung nach ist es', 9),
 (u'der meinung dass kriminalit\ex4t',
 9),
 (u'dass kriminalit\ex4t zahlt sich',
 9),
 (u'an der universit\ex4t zu', 9),
 (u'sagen dass kriminalit\ex4t sich',

9),
 (u'man sagen dass die', 9),
 (u'meisten universit\ex4tsabschl\xfcsse
 sind nicht', 9),
 (u'auf der andere seite', 9),
 (u'sich nicht aus weil', 9),
 (u'im großen und ganzen', 9),
 (u'aber meiner meinung nach', 9),
 (u'die rechte der frauen', 9),
 (u'f\xfcfr die gesellschaft leistet',
 9),
 (u'1 kriminalit\ex4t zahlt sich', 8),
 (u'es ist nicht nur', 8),
 (u'einen job zu finden', 8),
 (u'die frage ob die', 8),
 (u'um geld zu verdienen', 8),
 (u'von geringem wert weil', 8),
 (u'deswegen von geringen wert', 8),
 (u'f\xfcfr die gesellschaft zu', 8),
 (u'der anderen seite ist', 8),
 (u'die universit\ex4tsabschl\xfcsse
 nicht praxisorientiert', 8),
 (u'behaupten dass der feminismus', 8),
 (u'um einen job zu', 8),
 (u'die nicht praxisorientiert sind',
 8),
 (u'die wirkliche welt nicht', 8),
 (u'dass die
 universit\ex4tsabschl\xfcsse nicht',
 8),
 (u'zum schluss glaube ich', 7),
 (u'ich würde sagen dass', 7),
 (u'geld zu verdienen und', 7),
 (u'die rolle der frauen', 7),
 (u'vorbereitet auf die wirkliche', 7),
 (u'die universit\ex4tsabschl\xfcsse
 von geringem', 7),
 (u'schluss glaube ich dass', 7),
 (u'sind deswegen von geringen', 7),
 (u'in deutschland gibt es', 7),
 (u'in den meisten fällen', 7),
 (u'meinung nach sind die', 7),
 (u'drei oder vier jahren', 7),
 (u'wichtig f\xfcfr die gesellschaft',
 7),

(u'das bedeutet nicht dass', 7),
 (u'ob der feminismus den', 7),
 (u'zu hause zu bleiben', 7),
 (u'entsprechen den er f\xfcfr', 7),
 (u'in bezug auf die', 7),
 (u'der universit\ex4t zu studieren',
 7),
 (u'sich um die kinder', 7),
 (u'einen guten job zu', 7),
 (u'einen job zu bekommen', 7),
 (u'nicht zu leugnen dass', 7),
 (u'eine frau zu sein', 7),
 (u'die wirkliche welt zu', 6),
 (u'finanzielle entlohnung eines
 menschen', 6),
 (u'weil es so viele', 6),
 (u'in diesem aufsatz werde', 6),
 (u'sich nicht aus zahlt', 6),
 (u'interessen der frauen weil', 6),
 (u'gen\xfctzt als geschadet hat', 6),
 (u'kriminalit\ex4t zahlt sich aus',
 6),
 (u'von geringem wert sein', 6),
 (u'viel f\xfcfr die gesellschaft', 6),
 (u'in der wirkliche welt', 6),
 (u'diesem aufsatz werde ich', 6),
 (u'wirkliche welt vor weil', 6),
 (u'k\xf6nnte gesagt werden dass', 6),
 (u'es kommt darauf an', 6),
 (u'eine arbeit zu finden', 6),
 (u'nicht zu sagen dass', 6),
 (u'dass die
 universit\ex4tsabschl\xfcsse von', 6),
 (u'machen was sie wollen', 6),
 (u'wenn man z. b.', 6),
 (u'an der universit\ex4t studiert',
 6),
 (u'frauen die gleiche rechte', 6),
 (u'der anderen seite kann', 6),
 (u'dass feminismus den interessen',
 6),
 (u'ist meiner meinung nach', 6),
 (u'es ist möglich dass', 6),
 (u'nicht aus weil es', 6),
 (u'interessen der frauen nicht', 6),

(u'möchte ich sagen dass', 6),
 (u'frauen mehr geschadet hat', 6),
 (u'meinung dass der feminismus', 6),
 (u'zum schluss möchte ich', 6),
 (u'aber es gibt auch', 6),
 (u'ist es ganz anders', 6),
 (u'meiner meinung nach die', 6),
 (u'es gibt nicht so', 6),
 (u'sagen dass die kriminalit\xe4t',
 6),
 (u'zu sagen dass es', 6),
 (u'wenn man zum beispiel', 6),
 (u'im ausland zu studieren', 6),
 (u'es gibt so viele', 6),
 (u'dass m\xe4nner und frauen', 6),
 (u'dass die studenten nicht', 6),
 (u'ob es sich lohnt', 6),
 (u'universit\xe4tsabschl\xfcsses sind
 nicht praxisorientiert', 6),
 (u'meinung nach ist das', 6),
 (u'an der wirtschaftsuniversit\xe4t
 kopenhagen', 6),
 (u'die meisten
 universit\xe4tsabschl\xfcsses von', 6),
 (u'sehr wichtig f\xfcr die', 6),
 (u'die wirkliche welt sind', 6),
 (u'der meinung dass der', 6),
 (u'aber ich denke dass', 6),
 (u'der feminismus in der', 5),
 (u'entsprechen sollte den er', 5),
 (u'sich nicht auszahlt aber', 5),
 (u'besser auf die wirkliche', 5),
 (u'als gen\xfcgt der feminismus', 5),
 (u'geschadet als gen\xfcgt der', 5),
 (u'nichts f\xfcr die gesellschaft',
 5),
 (u'der feminismus ist eine', 5),
 (u'an der universit\xe4t gelernt', 5),
 (u'meisten
 universit\xe4tsabschl\xfcsses von
 geringem', 5),
 (u'es gibt auch viele', 5),
 (u'universit\xe4tsabschl\xfcsses die
 studenten auf', 5),
 (u'es nicht so einfach', 5),

(u'es ist aber nicht', 5),
 (u'man in der gesellschaft', 5),
 (u'hausa zu bleiben und', 5),
 (u'die wirkliche welt vorzubereiten',
 5),
 (u'es gibt viele gründe', 5),
 (u'einen beruf zu finden', 5),
 (u'dem beitrag entsprechen sollte',
 5),
 (u'um die kinder k\xfcmmern', 5),
 (u'frauen in der politik', 5),
 (u'muss man sagen dass', 5),
 (u'in der gesellschaft und', 5),
 (u'in der gesellschaft haben', 5),
 (u'feminismus die interessen der', 5),
 (u'die meisten
 universit\xe4tsabschl\xfcsses', 5),
 (u'zum beispiel in der', 5),
 (u'als m\xe4nner haben sollen', 5),
 (u'dass die leute die', 5),
 (u'es ist wichtig zu', 5),
 (u'der wirklichen welt zu', 5),
 (u'zum beispiel gibt es', 5),
 (u'wichtig zu sagen dass', 5),
 (u'der meinung dass es', 5),
 (u'den interesse der frauen', 5),
 (u'muss man sich fragen', 5),
 (u'die wirkliche welt ist', 5),
 (u'dass die frauen nicht', 5),
 (u'den beitrag eines menschen', 5),
 (u'beitrag entsprechen sollte den',
 5),
 (u'ist es wichtig dass', 5),
 (u'interessen der frauen schadet', 5),
 (u'es gibt leute die', 5),
 (u'von der wirklichen welt', 5),
 (u'es ist wahr dass', 5),
 (u'ein verbrechen zu begehen', 5),
 (u'sind und bereiten die', 5),
 (u'dass die meisten studenten', 5),
 (u'aber es ist auch', 5),
 (u'ein großes problem in', 5),
 (u'man darf nicht vergessen', 5),
 (u'die wirkliche welt und', 5),

(u'meiner meinung nach hat', 5),
 (u'f\xfcr mich ist es', 5),
 (u'ist sehr wichtig f\xfcr', 5),
 (u'anderen seite gibt es', 5),
 (u'sie von geringem wert', 5),
 (u'z. b. wenn man', 5),
 (u'kriminalit\xe4t ein gro\xdfes
 problem', 5),
 (u'es gibt viele andere', 5),
 (u'die situation der frauen', 5),
 (u'sind und dass sie', 5),
 (u'ist es auch sehr', 5),
 (u'studium an der universit\xe4t', 5),
 (u'sich nicht aus kriminalit\xe4t',
 5),
 (u'dass frauen die gleiche', 5),
 (u'die man in der', 5),
 (u'man ein verbrechen begeht', 5),
 (u'aber man muss auch', 5),
 (u'zu sagen dass der', 5),
 (u'nicht zu hause bleiben', 5),
 (u'es ist nicht zu', 5),
 (u'nicht praxisorientiert sind aber',
 5),
 (u'anderen seite ist es', 5),
 (u'ich glaube dass es', 5),
 (u'würde ich sagen dass', 5),
 (u'viel geld zu verdienen', 5),
 (u'auch wenn sie nicht', 5),
 (u'überall in der welt', 5),
 (u'der frauen gen\xfcgt hat', 5),
 (u'mehr f\xfcr die gesellschaft', 5),
 (u'dass es nicht so', 5),
 (u'aber es ist nicht', 5),
 (u'der feminismus den frauen', 5),
 (u'der feminismus ist ein', 5),
 (u'was man in der', 5),
 (u'dass die neuen studiengänge',
 5),
 (u'dass die studenten die', 5),
 (u'mit drogen zu tun', 5),
 (u'sind die meisten
 universit\xe4tsabschl\xfcsses', 5),
 (u'aber auf der anderen', 5),
 (u'ist es klar zu', 5),

(u'ich denke dass die', 5),

Liste 3. L2_n3_PMI

[(u'simone de beauvoir',	7.367467414715759),	6.799717158525262),
10.09391870118886),	(u'ins gef\xe4ngnis gehen',	(u'am wichtigsten ist',
(u'vorlesungen und seminaren',	7.3370783359172185),	6.794121433968434),
9.23646846933764),	(u'seit vielen jahren',	(u'ab und zu', 6.794121433968434),
(u'vor- und nachteile',	7.278723862635323),	(u'eines menschen dem',
9.226418133484136),	(u'z. b. in', 7.2473495659510885),	6.761554764947157),
(u'u. s. w.', 8.898142664067104),	(u'davon ab wie', 7.228864751276985),	(u'darf nicht vergessen',
(u'd. h. dass', 8.689924762969568),	(u'drogen zu kaufen',	6.7290889636975795),
(u'positive und negative',	7.191712485645692),	(u'geringem wert sein',
8.638631468582018),	(u'finanzielle entlohnung eines',	6.7119526814545925),
(u'art und weise', 8.302159231960806),	7.118617479542931),	(u'entlohnung eines menschen',
(u'bachelor und master',	(u'entsprechen sollte den',	6.688563309378041),
8.168627839336283),	7.116204933137547),	(u'entlohnung eines menschen',
(u'gro\xdfen und ganzen',	(u'zwischen den geschlechtern',	6.686843618498514),
8.0508448036799),	7.107155097617629),	(u'glaube ich dass',
(u'drei oder vier',	(u'hausa zu bleiben',	6.682190231905916),
8.032495665011703),	7.095066300298625),	(u'gleiche rechte wie',
(u'stunden pro woche',	(u'hausa bleiben und',	6.6444171820983104),
7.970802096006363),	7.066895423331929),	(u'geringem wert sind',
(u'darauf an wie', 7.850174108217748),	(u'bin ich der', 7.0429237489598195),	6.594169645798209),
(u'berufliche stelle zu',	(u'ihr eigenes geld',	(u'geringem wert weil',
7.80789862506951),	7.031696686366036),	6.594169645798209),
(u'laufe der zeit',	(u'verbrechen zu begehen',	(u'gleiche rechte haben',
7.773634031095414),	6.997210510294867),	6.5762089320717765),
(u'darauf an was', 7.732391072561365),	(u'ziele zu erreichen',	(u'diesem aufsatz werde',
(u'glauben viele dass',	6.980403391978485),	6.550891124087487),
7.673550572538431),	(u'zwischen m\xe4nnern und',	(u'alles in allem',
(u'recht zum w\xe4hlen',	6.953004417790371),	6.521171859714661),
7.664071828583887),	(u'davon \xfcbberzeugt dass',	(u'meinung nach hat',
(u'bis zum ende', 7.63432985938515),	6.946297779491975),	6.489780475530014),
(u'gro\xdfes problem in',	(u'sieht man dass', 6.92633617070821),	(u'meinung nach sind',
7.526773953163888),	(u'meiner meinung nach',	6.489780475530014),
(u'entsprechen den er',	6.891376851029132),	(u'm\xe4nnern und frauen',
7.496027610632374),	(u'beitrag entsprechen den',	6.485209513904191),
(u'vorteile und nachteile',	6.8604721277838125),	(u'beitrag entsprechen sollte',
7.475480658776338),	(u'geschadet als gen\xfctzt',	6.47100736102209),
(u'z. b. wenn', 7.429671122745043),	6.841141324644978),	(u'meinung nach ist',
(u'z. b. die', 7.429671122745043),	(u'ausland zu studieren',	6.4505597623767335),
(u'jahre alt sind',	6.839491892056341),	(u'dieses thema ist',
7.4266904946069054),	(u'gen\xfctzt als geschadet',	6.449775140916316),
(u'typen von kriminalit\xe4t',	6.83889111102295),	(u'zahlt sich aus',
7.383015431104244),	(u'eines menschen sollte',	6.445781525384845),
(u'unterschiede zwischen den',	6.822179386763592),	(u'aufsatz werde ich',
	(u'egal ob man', 6.81235557396985),	6.44317613219764),
	(u'gesagt werden dass',	(u'wirkliche welt vor',

6.440073715991495),
 (u'aufsatz wird die',
 6.4180702110665635),
 (u'am ende des', 6.406355902959671),
 (u'praxisorientiert und bereiten',
 6.356530167027171),
 (u'anderen seite ist',
 6.342630503464988),
 (u'meinung nach es',
 6.30745891873606),
 (u'meisten
 universit\xe4tsabschl\xfcsse sind',
 6.306892185935399),
 (u'wirkliche welt vorbereiten',
 6.305384915830361),
 (u'lohnt es sich',
 6.3012058709272525),
 (u'werde ich die', 6.295224381067717),
 (u'vor dem gericht',
 6.287256211418541),
 (u'vor allem die', 6.287256211418541),
 (u'geringem wert ist',
 6.265665578826173),
 (u'm\xfcchte ich sagen',
 6.243336977483705),
 (u'wirkliche welt vorbereitet',
 6.234027977048529),
 (u'guten job zu', 6.225591714293427),
 (u'aufsatz m\xfcchte ich',
 6.205895691122928),
 (u'zahlt sich nicht',
 6.195782886767459),
 (u'kommt darauf an',
 6.187010796427485),
 (u'sollte dem beitrag',
 6.174948455245644),
 (u'menschen sollte dem',
 6.165316566957803),
 (u'wirklichen welt zu',
 6.163494610389823),
 (u'denke ich dass',
 6.1443934913651335),
 (u'wirkliche welt vorzubereiten',
 6.1359667638723145),
 (u'kriminell zu werden',
 6.114153606175416),
 (u'schluss m\xfcchte ich',
 6.104273497575708),
 (u'w\xendend des studiums',
 6.100315708768291),
 (u'meinung nach die',
 6.0843153674218495),
 (u'interessen der frauen',
 6.0780942460545235),
 (u'lernt man viele',
 6.075221757306074),
 (u'l kriminalit\xe4t zahlt',
 6.0612595911219245),
 (u'meisten universit\xe4tsabschl\xfcsse
 sind', 6.055577757654493),
 (u'anderen seite kann',
 6.054948431013207),
 (u'sowohl als auch',
 6.052748901454016),
 (u'gleich wie m\xe4nner',
 6.036682798362354),
 (u'deswegen von geringen',
 6.030862113937553),
 (u'nach dem universit\xe4tsabschluss',
 5.996009072802264),
 (u'bei einer firma',
 5.981874561867358),
 (u'bereiten die studenten',
 5.962747224796983),
 (u'universit\xe4tsabschl\xfcsse sind
 nicht', 5.956005918778479),
 (u'am anfang des', 5.946823573581231),
 (u'des studiums zu',
 5.940631603333309),
 (u'beitrag zur gesellschaft',
 5.938202830537324),
 (u'deswegen von geringem',
 5.9365514344663115),
 (u'fr\xcher war es',
 5.934589288877347),
 (u'gab es eine', 5.924350717729172),
 (u'schluss glaube ich',
 5.91358369111237),
 (u'bin der meinung',
 5.911777686631557),
 (u'gab es die', 5.896179840762476),
 (u'vor 50 jahren', 5.881791103310376),
 (u'anderen seite gibt',
 5.872626874219252),
 (u'frage zu beantworten',
 5.872312359355832),
 (u'dieser aufsatz wird',
 5.864091526210975),
 (u'nach dem abschluss',
 5.8529082291615895),
 (u'zum beispiel gibt',
 5.838305991370637),
 (u'zum schluss ist',
 5.838305991370637),
 (u'zum beispiel mit',
 5.838305991370637),
 (u'zum ersten mal',
 5.838305991370637),
 (u'zum beispiel im',
 5.838305991370637),
 (u'zum beispiel haben',
 5.838305991370637),
 (u'zum schluss glaube',
 5.838305991370637),
 (u'zum beispiel viele',
 5.838305991370637),
 (u'zum beispiel hat',
 5.838305991370637),
 (u'zum beispiel die',
 5.838305991370637),
 (u'zum beispiel medizin',
 5.838305991370637),
 (u'zum schluss m\xfcchte',
 5.838305991370637),
 (u'zum beispiel wie',
 5.838305991370637),
 (u'universit\xe4tsabschl\xfcsse sind
 nicht', 5.8224745261539566),
 (u'fragen ob die', 5.796633294970069),
 (u'kinder zu k\xfcmmern',
 5.795176630473605),
 (u'teil des lebens',
 5.78985360798469),
 (u'des 20. jahrhunderts',
 5.786480923506051),

(u'einen guten job',
 5.784894880350416),
 (u'zum beispiel in',
 5.7842387701003615),
 (u'weniger geld als',
 5.7835990771311625),
 (u'dem zweiten weltkrieg',
 5.779291819933582),
 (u'bei der arbeit',
 5.777673595892146),
 (u'an der wirtschaftsuniversit\xe4t',
 5.742324975166039),
 (u'zum beispiel wenn',
 5.738222532813655),
 (u'aus diesem grund',
 5.73480891296173),
 (u'zum beispiel ein',
 5.713142848416632),
 (u'k\xf6nnte man sagen',
 5.69744858654042),
 (u'dem beitrag entsprechen',
 5.693524998176157),
 (u'meisten
 universit\xe4tsabschl\xfcsse
 praxisorientiert', 5.672585505398388),
 (u'klar zu sehen', 5.672070572328307),
 (u'problem zu l\xf6sen',
 5.656998119407529),
 (u'im rahmen der', 5.646656243905739),
 (u'im laufe des', 5.646656243905739),
 (u'wirkliche welt sind',
 5.645760427309059),
 (u'entscheiden ob sie',
 5.6280105825342766),
 (u'gen\xfctzt hat weil',
 5.6255505566934145),
 (u'wirkliche welt nicht',
 5.625141140106324),
 (u'kein geld haben',
 5.623096567731848),
 (u'immer mehr studenten',
 5.6117235775595775),
 (u'sondern auch f\xfcr',
 5.605951362056373),
 (u'finde ich dass',
 5.603438126147114),
 (u'er sie f\xfcr', 5.591085202893747),
 (u'welt vor weil', 5.589654680509215),
 (u'zum schluss kann',
 5.586991563089731),
 (u'job zu kriegen',
 5.583216192866853),
 (u'im vergleich mit',
 5.582117722768168),
 (u'uns fragen ob', 5.571509198941354),
 (u'dies bedeutet dass',
 5.5585580465942765),
 (u'so viel wie', 5.547589015223702),
 (u'einen schlechten ruf',
 5.543732823533528),
 (u'bedeutet nicht dass',
 5.541923277688025),
 (u'im gro\xdfen und',
 5.541295728247913),
 (u'nach dem studium',
 5.53064582311303),
 (u'geld zu verdienen',
 5.5284550606371585),
 (u'wir m\xfcsen uns',
 5.523213209225558),
 (u'k\xf6nnte gesagt werden',
 5.510410011981948),
 (u'w\xfcrde ich sagen',
 5.509923512337227),
 (u'lohnt oder nicht',
 5.49814971114183),
 (u'im gef\xe4hnis sein',
 5.492505564078481),
 (u'im laufe der', 5.492505564078481),
 (u'vergleich mit den',
 5.485895446218362),
 (u'praktikum zu machen',
 5.478385728165223),
 (u'im vergleich zu',
 5.472302856760962),
 (u'studiert hat und',
 5.467546307550165),
 (u'immer mehr leute',
 5.4575728977323195),
 (u'mehr gen\xfctzt als',
 5.45331242724266),
 (u'kriminalit\xe4t zahlt sich',
 5.451064773181018),
 (u'mehr geschadet als',
 5.4290197346736155),
 (u'wirkliche welt und',
 5.42261687599485),
 (u'seite haben wir',
 5.420155723735157),
 (u'gesellschaft geleistet hat',
 5.416843222266782),
 (u'nach dem ersten',
 5.408222407900144),
 (u'kriminell zu sein',
 5.399392949194933),
 (u'm\xe4nner und frauen',
 5.392049991226319),
 (u'rechte wie m\xe4nner',
 5.386858299789707),
 (u'zum beispiel der',
 5.38632086762758),
 (u'kriminalit\xe4t zahlt sich aus',
 5.371600601826771),
 (u'mehr oder weniger',
 5.3580022474383355),
 (u'ende des studiums',
 5.355467153831455),
 (u'an der uni', 5.345954596963757),
 (u'feminismus hat den',
 5.3410738616796225),
 (u'auszahlt oder nicht',
 5.3312869957262095),
 (u'zum beispiel ist',
 5.327480367604647),
 (u'vorbereitung auf die',
 5.323117674483485),
 (u'auf jeden fall',
 5.315550080471123),
 (u'erfolgreich zu sein',
 5.288727381307413),
 (u'drogen zu tun', 5.286624331110635),
 (u'tatsache ist dass',
 5.279993701338658),
 (u'diese f\xe4higkeiten sind',
 5.275655299740061),

(u'meisten	5.184947060869105),	(u'meine meinung nach',
universit\xe4tsabschl\xfcsse nicht',	(u'an den universit\xe4ten',	5.013415918194992),
5.271458798888821),	5.182709187230616),	(u'auf jedem fall',
(u'mehr als nur', 5.270990870448705),	(u'auf keinen fall',	5.007248720816607),
(u'studenten nicht auf',	5.176666872774653),	(u'ich bin davon', 4.996450617477322),
5.268355541288191),	(u'viele leute glauben',	(u'ich denke nicht',
(u'obwohl es oft', 5.252886191478784),	5.156366263831081),	4.996450617477322),
(u'behaupten dass der',	(u'universit\xe4tsabschl\xfcsse nicht	(u'ich bin auch', 4.996450617477322),
5.252359737136539),	praxisorientiert', 5.14338733116231),	(u'ich meine dass',
(u'gibt viele gr\xfcnde',	(u'immer mehr frauen',	4.996450617477322),
5.249506668154132),	5.137265597964461),	(u'ich hoffe dass',
(u'ihre rechte und',	(u'wie schon erw\xe4hnt',	4.996450617477322),
5.243452159245425),	5.120985274276616),	(u'sehr wichtig dass',
(u'zwischen m\xe4nnern und',	(u'job zu bekommen',	4.989597923015281),
5.242370493374684),	5.118910584735755),	(u'wirkliche welt zu',
(u'im ausland zu', 5.241191135797575),	(u'mehr wert als', 5.116840190621447),	4.984361945063695),
(u'praxisorientiert sind oder',	(u'bin ich nicht', 5.109334084292547),	(u'den ganzen tag',
5.237267434140541),	(u'auf die wirkliche',	4.978923391768362),
(u'an der universit\xe4t',	5.089143883522097),	(u'wenn man z.', 4.96624138666094),
5.228171561307374),	(u'erwartet dass man',	(u'an der universit\xe4ten',
(u'so schnell wie',	5.083521530679871),	4.964019986265787),
5.211116778602489),	(u'm\xe4nnern haben sollen',	(u'so einfach ist',
(u'auf das arbeitsleben',	5.080837421364344),	4.959802350321583),
5.210189564813297),	(u'2 die meisten', 5.078888317758382),	(u'gemacht hat und',
(u'wegen des feminismus',	(u'so wichtig wie',	4.955267195156192),
5.209744094790517),	5.077585385977967),	(u'sehr schwierig zu',
(u'an einer universit\xe4t',	(u'heutzutage haben frauen',	4.950377209862),
5.205182043082675),	5.072979531954332),	(u'ich bin der', 4.945157323089771),
(u'was passiert wenn',	(u'auf der anderen',	(u'ich denke dass',
5.2046442640968715),	5.066734174102787),	4.942383396207046),
(u'job zu finden',	(u'nach ihrem studium',	(u'm\xfcssen wir uns',
5.1975537120548685),	5.051547463961412),	4.940668395735039),
(u'wirkliche welt ist',	(u'andererseits gibt es',	(u'auf dem arbeitsmarkt',
5.1915051550314635),	5.042461412572349),	4.938255849329655),
(u'dem ersten weltkrieg',	(u'als je zuvor', 5.039950282386231),	(u'frage ob die', 4.935402340278942),
5.1915051550314635),	(u'als gen\xfcgt der',	(u'er oder sie', 4.929813230864251),
(u'meisten	5.03995028238623),	(u'jetzt k\xfcnnen wir',
universit\xe4tsabschl\xfcsse von',	(u'auf der stra\xdfen',	4.929459311200319),
5.18807718994977),	5.038339307886638),	(u'anders als die',
(u'einen beruf zu',	(u'auf der einen', 5.03356602913398),	4.927379163616356),
5.1870578795947955),	(u'schwer zu sagen',	(u'gibt es andere',
(u'geworden ist und',	5.028795221808535),	4.921002601182096),
5.185157942354492),	(u'interesse der frauen',	(u'zeit im gef\xe4hrnis',
(u'mit einem	5.028628123470774),	4.918450193331986),
universit\xe4tsabschluss',	(u'einen job zu', 5.02154344117222),	(u'sondern auch in',

4.917423127352722),	4.845391999143839),	4.716392820964216),
(u'von geringerem wert',	(u'den meisten f\xe4llen',	(u'frauen mehr gen\xftzt',
4.916559047112318),	4.845391999143839),	4.713889690746098),
(u'von geringen wert',	(u'auf das wirkliche',	(u'was sie wollen',
4.916559047112318),	4.8328953336718286),	4.7118371458335595),
(u'ich glaube dass',	(u'ich finde dass',	(u'ich w\xfcrde sagen',
4.911292809137015),	4.814129060683367),	4.708768545025541),
(u'besser auf das',	(u'ich bin nicht', 4.814129060683367),	(u'wir sind alle', 4.701233156822644),
4.905735129051307),	(u'ich weiss nicht',	(u'sind deswegen von',
(u'von geringem wert',	4.814129060683367),	4.69811582224059),
4.891743877992594),	(u'sondern auch eine',	(u'bezug auf die', 4.695109796094501),
(u'ich finde es', 4.891090101819496),	4.811281666711824),	(u'ein gro\xdfes problem',
(u'was man machen',	(u'besser als m\enne',	4.694060605085863),
4.886190532978337),	4.808632298632463),	(u'ein verbrechen begangen',
(u'sich lohnt oder',	(u'den interessen der',	4.694060605085863),
4.884666767812147),	4.803635861631787),	(u'ein besseres leben',
(u'praxisorientiert sind und',	(u'sich nicht aus', 4.79505460912246),	4.694060605085863),
4.883693267880058),	(u'zuerst muss man',	(u'frauen und m\enne',
(u'keine zeit f\xfcr',	4.789122368543114),	4.6938708238556055)]
4.88115922300247),	(u'eine gute idee',	
(u'4 die meisten', 4.879759642648047),	4.7852974595025515),	
(u'sondern auch die',	(u'eine sehr gro\xdf',	
4.879104263050586),	4.7852974595025515),	
(u'kriminalit\et sich auszahlen',	(u'um ihre kinder',	
4.870825313914281),	4.769705771511263),	
(u'viele leute dass',	(u'den er sie', 4.769202860786293),	
4.8686841913793),	(u'auf der andere',	
(u'l\essst sich nicht',	4.7683568125342575),	
4.868172027475659),	(u'mehr verdienen als',	
(u'gibt es viele', 4.867298422210235),	4.760165246682715),	
(u'sich nicht auszahlt',	(u'f\higkeiten die man',	
4.864047480609412),	4.754596499694041),	
(u'meine ich dass',	(u'mit dem geld', 4.751311075794244),	
4.863459645903069),	(u'sehr wichtig in',	
(u'auf den arbeitsmarkt',	4.742737845083756),	
4.858213141656119),	(u'nur ein kleines',	
(u'rechte als m\enne',	4.737669410513297),	
4.857058892537937),	(u'als gen\xftzt hat',	
(u'ein gutes beispiel',	4.736267868588008),	
4.848211284913121),	(u'viel weniger als',	
(u'ein halbes jahr',	4.725276017920093),	
4.848211284913121),	(u'ob es sich', 4.725154974525889),	
(u'ein bisschen anders',	(u'mit den kindern',	
4.848211284913121),	4.72005853229014),	
(u'den letzten jahren',	(u'frauen mehr geschadet',	

Liste 4. L2_n4_PMI

[(u'drei oder vier jahren',	6.838772475203532),	(u'wirkliche welt vor weil',
8.142179176517654),	(u'eines menschen sollte dem',	6.189079923889872),
(u'entsprechen sollte den er',	6.826502382611717),	(u'menschen sollte dem beitrags',
7.726663732555989),	(u'geschadet als gen\xfctzt hat',	6.169639562805929),
(u'hausa zu bleiben und',	6.782037528208223),	(u'ob es sich lohnt',
7.659005084082174),	(u'eines menschen dem beitrags',	6.145625294643587),
(u'entsprechen den er sie',	6.761963861474146),	(u'l kriminalit\xe4t zahlt sich',
7.583562888915315),	(u'beitrags entsprechen sollte den',	6.145625294643586),
(u'z. b. wenn man',	6.698473908184425),	(u'deswegen von geringen wert',
7.433994118593168),	(u'interessen der frauen weil',	6.035185109785679),
(u'entsprechen den er f\xfcr',	6.582440932635827),	(u'bin der meinung dass',
7.274678608812931),	(u'interessen der frauen schadet',	6.027277583458949),
(u'finanzielle entlohnung eines	6.582440932635827),	(u'interessen der frauen nicht',
menschen', 7.2258884446435),	(u'diesem aufsatz werde ich',	5.976305129065511),
(u'finanzielle entlohnung eines	6.555214119935613),	(u'deswegen von geringem wert',
menschen', 7.124535950383212),	(u'interessen der frauen mehr',	5.936918302155221),
(u'bin ich der meinung',	6.526087996084695),	(u'einen guten job zu',
7.121354716961667),	(u'meinung nach ist das',	5.884528056002867),
(u'meiner meinung nach die',	6.494103471378141),	(u'anderen seite gibt es',
7.069143729639195),	(u'meinung nach ist es',	5.876949870067378),
(u'bereiten die studenten nicht',	6.49410347137814),	(u'zum schluss glaube ich',
7.034997648277842),	(u'meinung nach sind die',	5.842628987218763),
(u'gen\xftzt als geschadet hat',	6.49410347137814),	(u'zum beispiel in der',
6.957327413638497),	(u'interessen der frauen geschadet',	5.842628987218763),
(u'zwischen m\xe4nnern und frauen',	6.4914691544301),	(u'zum beispiel gibt es',
6.957327413638497),	(u'anderen seite ist es',	5.842628987218763),
(u'zahlt sich nicht aus',	6.483085673637693),	(u'zum beispiel wenn man',
6.935612337014672),	(u'm\xftchte ich sagen dass',	5.842628987218763),
(u'geschadet als gen\xftzt der',	6.468202742945983),	(u'zum schluss m\xftchte ich',
6.92018786668904),	(u'schluss glaube ich dass',	5.842628987218763),
(u'meiner meinung nach ist',	6.45690318769205),	(u'an der wirtschaftsuniversit\xe4t
6.894790342494417),	(u'sollte dem beitrags entsprechen',	kopenhagen', 5.815640842501116),
(u'meiner meinung nach hat',	6.443012494117425),	(u'er sie f\xftcr die',
6.88682217284524),	(u'meisten	5.800852172956682),
(u'beitrags entsprechen den er',	universit\xe4tsabschl\xftsse sind	(u'k\xftnte man sagen dass',
6.8647951236319384),	nicht', 6.410306084427756),	5.774244203315159),
(u'entlohnung eines menschen dem',	(u'interessen der frauen gen\xftzt',	(u'geld zu verdienen und',
6.847036985053425),	6.374801567857582),	5.766971443986079),
(u'entlohnung eines menschen sollte',	(u'praxisorientiert und bereiten die',	(u'zwischen m\xe4nner und frauen',
6.847036985053425),	6.367727392046867),	5.748173250570564),
(u'meiner meinung nach sind',	(u'meisten universit\xe4tsabschl\xftsse	(u'dem beitrags entsprechen den',
6.846000178324985),	sind nicht', 6.360005345952957),	5.734017874642336),
(u'\xfberall in der welt',	(u'universit\xe4tsabschl\xftsse sind	(u'universit\xe4tsabschl\xftsse nicht
	nicht praxisorientiert',	praxisorientiert sind',
	6.282345242289056),	5.717574681879144),

(u'universit\xe4tsabschl\xfcscse sind nicht praxisorientiert',	(u'um die kinder k\xfcfcmmern',	(u'von geringem wert sein',
5.69304800308157),	5.371865768115009),	4.920882042960444),
(u'k\xfcfnnte gesagt werden dass',	(u'an der universit\xe4t gelernt',	(u'von geringem wert weil',
5.677251937327849),	5.34563721325538),	4.920882042960444),
(u'behaupten dass der feminismus',	(u'mit drogen zu tun',	(u'sagen dass kriminalit\xe4t sich',
5.661701512140987),	5.343420736544489),	4.90251091926405),
(u'an der universit\xe4t studiert',	(u'w\xfcfrde ich sagen dass',	(u'einen job zu bekommen',
5.661490162673858),	5.301153292724645),	4.88599922589174),
(u'im gro\xdcdfen und ganzen',	(u'auf der anderen seite',	(u'von geringem wert sind',
5.650979239753865),	5.266365922271678),	4.881661329807162),
(u'an der universit\xe4t studieren',	(u'im ausland zu studieren',	(u'ich w\xfcfrde sagen dass',
5.64858675783795),	5.245514131645701),	4.867242220700925),
(u'dem beitrag entsprechen sollte',	(u'auf der andere seite',	(u'ein gro\xdcdfes problem in',
5.601293258987753),	5.242683437628119),	4.852534280761247),
(u'2 die meisten universit\xe4tsabschl\xfcscse',	(u'meisten universit\xe4tsabschl\xfcscse von geringem',	(u'den interessen der frauen',
5.594036937372499),	5.241590429988667),	4.84141619217727),
(u'kriminalit\xe4t zahlt sich nicht',	(u'an der universit\xe4t zu',	(u'von geringem wert ist',
5.583562962453141),	5.2402766975975545),	4.833870665970814),
(u'4 die meisten universit\xe4tsabschl\xfcscse',	(u'auf der einen seite',	(u'ich denke dass die',
5.577229819056118),	5.180989868622779),	4.818452056531493),
(u'kriminalit\xe4t sich nicht auszahlt',	(u'rolle in der gesellschaft',	(u'wichtig zu sagen dass',
5.5582451544688505),	5.144554754811626),	4.812901021869249),
(u'vorbereitung auf die wirkliche',	(u'sagen dass die kriminalit\xe4t',	(u'den er f\xfcfr die',
5.551230425224265),	5.143672976080938),	4.800924830822533),
(u'einen beruf zu finden',	(u'einen job zu finden',	(u'f\xfcfr die gesellschaft machen',
5.548055819381654),	5.130756253626487),	4.787501810490392),
(u'kriminalit\xe4t zahlt sich aus',	(u'auf die wirkliche welt',	(u'f\xfcfr mich ist es',
5.509454990299419),	5.107185894039418),	4.787501810490392),
(u'meisten universit\xe4tsabschl\xfcscse nicht praxisorientiert',	(u'sich nicht auszahlt aber',	(u'f\xfcfr die gesellschaft leistet',
5.49596530442268),	5.071311320454228),	4.787501810490392),
(u'muss man sich fragen',	(u'sich nicht aus zahlt',	(u'sehr wichtig f\xfcfr die',
5.490218442066488),	(u'als gen\xfcftzt der feminismus',	4.757532140799177),
(u'mehr gen\xfcftzt als geschadet',	5.044273278234357),	(u'f\xfcfr die gesellschaft geleistet',
5.457635423090786),	(u'ich glaube dass die',	4.747227911352452),
(u'mehr geschadet als gen\xfcftzt',	5.000773613325448),	(u'frauen mehr geschadet als',
5.431433050696762),	(u'ich glaube dass es',	4.740904097796293),
(u'feminismus hat den interessen',	5.000773613325448),	(u'sind deswegen von geringem',
5.424600052069566),	(u'den er sie f\xfcfr',	4.739770407669202),
(u'studenten nicht auf die',	4.983246387616488),	(u'universit\xe4tsabschl\xfcscse die studenten auf',
5.375737558058635),	(u'wenn man z. b.',	4.73624553825767),
	4.970564382509066),	(u'sich nicht aus kriminalit\xe4t',
	(u'ich bin der meinung',	4.734839083833015),
	4.9436151994854995),	(u'sagen dass der feminismus',

4.73216553516811),
(u'sind nicht praxisorientiert und',
4.716966918651625),
(u'frauen mehr gen\xfctzt als',
4.713758336244843),
(u'als m\exenner haben sollen',
4.707801041613144),
(u'eine frau zu sein',
4.656089062726155),
(u'das leben der frauen',
4.650273555984543),
(u'das bedeutet nicht dass',
4.650273555984543),
(u'den beitrag eines menschen',
4.646774150995275),
(u'sich nicht aus weil',
4.629478568175188),
(u'sind deswegen von geringen',
4.624257520547357),
(u'f\xfcr die gesellschaft ist',
4.605180253696438),
(u'machen was sie wollen',
4.5673256204039845),
(u'weil es so viele',
4.566264837446764),
(u'beitrag f\xfcr die gesellschaft',
4.5033090638112165),
(u'frauen mehr geschadet hat',
4.495069135280013),
(u'ein praktikum zu machen',
4.447069172653082),
(u'menschen dem beitrag entsprechen',
4.42471761993142),
(u'um einen job zu',
4.414753041720599),
(u'ob der feminismus den',
4.4026559895849635),
(u'kriminalit\exet ein gro\xdfes
problem', 4.382671824664612),
(u'ein verbrechen zu begehen',
4.382530651515511),
(u'sie sind deswegen von',
4.355697428241984),
(u'man darf nicht vergessen',
4.345669058211353),

(u'kann man sagen dass',
4.339855088345933),
(u'hat den interessen der',
4.248912672851151),
(u'sich um die kinder',
4.24463274726976),
(u'in diesem aufsatz werde',
4.224113856633546),
(u'in deutschland gibt es',
4.224113856633545),
(u'in den meisten f\exellen',
4.224113856633545),
(u'in den letzten jahren',
4.224113856633545),
(u'feminismus den interessen der',
4.221322969085612),
(u'man k\xfennnte sagen dass',
4.21213766558683),
(u'um geld zu verdienen',
4.167892963789073),
(u'nicht zu leugnen dass',
4.165617649951938),
(u'es gibt auch viele',
4.15769457511461),
(u'es kann sein dass',
4.15769457511461),
(u'es gibt viele gr\xfende',
4.15769457511461),
(u'es ist wichtig zu',
4.15769457511461),
(u'es ist m\xfglich dass',
4.15769457511461),
(u'es gibt nicht so',
4.15769457511461),
(u'es gibt leute die',
4.15769457511461),
(u'es gibt so viele',
4.15769457511461),
(u'es gibt viele andere',
4.15769457511461),
(u'ist es wichtig dass',
4.156286322523053),
(u'zu hause bleiben und',
4.150952976547691),
(u'in der heutigen gesellschaft',

4.150005884479824),
(u'nicht praxisorientiert und
bereiten', 4.131131473880768),
(u'aber man muss auch',
4.098171428457148),
(u'vorbereitet auf die wirkliche',
4.095943192615585),
(u'es ist klar dass',
4.083586602960889),
(u'universit\exetsabschl\xfcsse von
geringem wert', 4.034011990709864),
(u'f\xfcr die gesellschaft zu',
4.033730008114012),
(u'zu hause zu bleiben',
4.017421583923168),
(u'nicht aus weil es',
4.011466970124679),
(u'man ein verbrechen begeht',
4.00919682159014),
(u'es kommt darauf an',
4.003543895287351),
(u'frauen in der politik',
3.994293847367524),
(u'muss man sagen dass',
3.986141045290214),
(u'es ist wahr dass',
3.9753730183206555),
(u'ist es auch sehr',
3.9739647657290984),
(u'ist es klar zu',
3.9739647657290984),
(u'in bezug auf die',
3.9727994283526398),
(u'dass die
universit\exetsabschl\xfcsse nicht',
3.9408844537020933),
(u'frauen die gleiche rechte',
3.9354533473445907),
(u'nicht praxisorientiert sind und',
3.929228871887707),
(u'wenn man zum beispiel',
3.9207422580103883),
(u'in der gesellschaft zu',
3.9139589283297065),
(u'aber ich denke dass',

3.875027877142938),	3.7193911552207455),	(u'es ist nicht nur',
(u' der anderen seite kann',	(u'feminismus die interessen der',	3.40392277273823),
3.873541835048004),	3.7185672023817786),	(u' der feminismus ist ein',
(u' der meinung dass kriminalit\xe4t',	(u'dass kriminalit\xe4t sich nicht',	3.4035382058022683),
3.8735418350480035),	3.7044956756378626),	(u'dass frauen die gleiche',
(u' der anderen seite ist',	(u'nicht auf die wirkliche',	3.390838116782821),
3.8735418350480035),	3.7039822703767182),	(u'dass die neuen studieng\xe4nge',
(u' der frauen gen\xfcgt hat',	(u' der wirklichen welt zu',	3.390838116782821),
3.8735418350480035),	3.691220278254049),	(u' hat der feminismus den',
(u' ist es klar dass',	(u' studium an der universit\xe4t',	3.343728419945889),
3.868604250071272),	3.640351342031152),	(u'dass es nicht so',
(u' ist es ganz anders',	(u' und bereiten die studenten',	3.3167301446290995),
3.868604250071272),	3.639991191839256),	(u' ich der meinung dass',
(u' meinung dass der feminismus',	(u' eine arbeit zu finden',	3.3043243239017177),
3.8670223328095976),	3.636940945412292),	(u' aber es ist nicht',
(u' viel geld zu verdienen',	(u' der feminismus den interessen',	3.250873568069944),
3.854130276414319),	3.6351308116030054),	(u' kriminalit\xe4t sich nicht aus',
(u' der frauen mehr geschadet',	(u'dass die kriminalit\xe4t sich',	3.247691891825628),
3.8327198405277487),	3.6204125584273217),	(u' die wirkliche welt vorbereiten',
(u' nicht praxisorientiert sind aber',	(u' in der wirkliche welt',	3.2459236249460743),
3.8291454133307243),	3.6179780530632306),	(u' die wirkliche welt vorzubereiten',
(u' was man in der',	(u' sind und dass sie',	3.2459236249460743),
3.822672898825107),	3.5946381033661994),	(u' die frage ob die',
(u' auch wenn sie nicht',	(u' der feminismus in der',	3.2459236249460743),
3.8136811876331427),	3.537069598426791),	(u' die wirkliche welt ist',
(u' der feminismus hat den',	(u' der feminismus ist eine',	3.2459236249460743),
3.8116664313299164),	3.537069598426791),	(u' die situation der frauen',
(u' der meinung dass die',	(u' der frauen geschadet hat',	3.2459236249460743),
3.8090033139104325),	3.537069598426791),	(u' die meisten
(u' dass der feminismus den',	(u' der meinung dass es',	universit\xe4tsabschl\xfcsse sind',
3.8073530610775705),	3.537069598426791),	3.2459236249460743),
(u' den interesse der frauen',	(u' der feminismus den frauen',	(u' die gesellschaft geleistet hat',
3.7594709559943715),	3.537069598426791),	3.231937382971335),
(u' in der wirklichen welt',	(u' der universit\xe4t zu studieren',	(u' sind der meinung dass',
3.757090855525948),	3.5168668911092715),	3.228393708411317),
(u' in der gesellschaft haben',	(u' wichtig f\xfcr die gesellschaft',	(u' die wirkliche welt vor',
3.75411022738781),	3.492433012071358),	3.225989410045257),
(u' dass die	(u' es ist aber nicht',	(u' die studenten nicht auf',
universit\xe4tsabschl\xfcsse von',	3.464547394554665),	3.225016940126761),
3.740213758239942),	(u' es ist nicht zu',	(u' sie f\xfcr die gesellschaft',
(u' nichts f\xfcr die gesellschaft',	3.464547394554665),	3.2231369648629085),
3.731771249305548),	(u' ist sehr wichtig f\xfcr',	(u' die finanzielle entlohnung eines',
(u' der frauen mehr gen\xfcgt',	3.4631391419631075),	3.221532171821915),
3.7304409914073307),	(u' in der gesellschaft und',	(u' die meisten
(u' der meinung dass der',	3.4356564962692757),	universit\xe4tsabschl\xfcsse sind',

3.2177527479793784),	(u'der frauen in der',	1.9484542890118872),
(u'aber es ist auch',	2.7475305721917795),	(u'nicht von geringem wert',
3.215782248258674),	(u'aber auf der anderen',	1.868045098781436),
(u'er f\xfcr die gesellschaft',	2.7329304765350892),	(u'dass kriminalit\xe4t zahlt sich',
3.2120566781223117),	(u'zu sagen dass der',	1.8614429120222573),
(u'nicht so viel geld',	2.715868451258368),	(u'mehr f\xfcr die gesellschaft',
3.210106204924501),	(u'die interesse der frauen',	1.799215176461558),
(u'es nicht so einfach',	2.715295373883904),	(u'sind und bereiten die',
3.2021831300871737),	(u'die kriminalit\xe4t sich auszahlt',	1.5074144212443086),
(u'die meisten	2.6863078370106517),	(u'die kriminalit\xe4t sich nicht',
universit\xe4tsabschl\xfcsse nicht',	(u'dass die finanzielle entlohnung',	1.50545745010557),
3.1718156527923527),	2.523818433915449),	(u'die nicht praxisorientiert sind',
(u'die wirkliche welt vorbereitet',	(u'dass die studenten die',	1.4133421611977643),
3.165880917272538),	2.5218002697592117),	(u'sind die meisten
(u'studenten auf die wirkliche',	(u'zu sagen dass es',	universit\xe4tsabschl\xfcsse',
3.140463248115785),	2.5087252412905996),	1.4043821953460736),
(u'die wirkliche welt nicht',	(u'nicht zu hause bleiben',	(u'die man in der',
3.128140589289691),	2.442851052210834),	1.3896256345804483),
(u'dass frauen und m\xe4nner',	(u'dass die leute die',	(u'man in der gesellschaft',
3.122574130188142),	2.429426949628197),	1.2101748422822032),
(u'die wirkliche welt zu',	(u'dass m\xe4nner und frauen',	(u'in die wirkliche welt',
3.0917729451188163),	2.37690891534475),	1.0886196407043962),
(u'die meisten	(u'die universit\xe4tsabschl\xfcsse	(u'dass die studenten nicht',
universit\xe4tsabschl\xfcsse von',	von geringem', 2.3586204299451716),	1.0818632256900045),
3.0917729451188163),	(u'nicht zu sagen dass',	(u'ist meiner meinung nach',
(u'die wirkliche welt sind',	2.346459206535768),	1.0652438691647372),
3.0917729451188163),	(u'frauen in der gesellschaft',	(u'die interessen der frauen',
(u'besser auf die wirkliche',	2.34040417236259),	0.9350397291373337),
3.0753736109543444),	(u'dass die meisten	(u'sie von geringem wert',
(u'dass die meisten studenten',	universit\xe4tsabschl\xfcsse',	0.6799938265907531)]
3.0654157163481934),	2.3279439107167907),	
(u'aber es gibt auch',	(u'von der wirklichen welt',	
3.0518275683347946),	2.3256273360035777),	
(u'dass die frauen nicht',	(u'aber meiner meinung nach',	
3.0113484950779172),	2.288062820560896),	
(u'die rechte der frauen',	(u'viel f\xfcr die gesellschaft',	
2.9582415524942935),	2.2343295769446727),	
(u'die wirkliche welt und',	(u'man an der universit\xe4t',	
2.9094513883248614),	2.179216139541887),	
(u'die studenten auf die',	(u'f\xfcr die wirkliche welt',	
2.902978873819244),	2.0827905106536964),	
(u'die rolle der frauen',	(u'die universit\xe4tsabschl\xfcsse	
2.889248681007342),	nicht praxisorientiert',	
(u'man sagen dass die',	1.9580693366394364),	
2.8052240172642042),	(u'dass feminismus den interessen',	

Liste 5. L1_n3_mostcommon

(u'f\xfor die gesellschaft', 65),
 (u'meiner meinung nach', 36),
 (u'kriminalit\xe4t zahlt sich', 32),
 (u'zahlt sich nicht', 27),
 (u'sich nicht aus', 27),
 (u'die finanzielle entlohnung', 24),
 (u'sie f\xfor die', 21),
 (u'in den letzten', 19),
 (u'die gesellschaft geleistet', 18),
 (u'entlohnung eines menschen', 17),
 (u'dass sich kriminalit\xe4t', 17),
 (u'finanzielle entlohnung eines', 16),
 (u'mehr geschadet als', 16),
 (u'der meinung dass', 16),
 (u'interessen der frauen', 16),
 (u'gesellschaft geleistet hat', 16),
 (u'er sie f\xfor', 15),
 (u'den er sie', 15),
 (u'in den meisten', 15),
 (u'ich denke dass', 14),
 (u'frauen mehr geschadet', 14),
 (u'dem beitrag entsprechen', 14),
 (u'in unserer gesellschaft', 14),
 (u'sich kriminalit\xe4t nicht', 14),
 (u'in der gesellschaft', 14),
 (u'geschadet als gen\xftzt', 14),
 (u'den ganzen tag', 14),
 (u'den letzten jahren', 13),
 (u'in der heutigen', 13),
 (u'wie zum beispiel', 12),
 (u'der frauen mehr', 12),
 (u'beitrag entsprechen den', 12),
 (u'entsprechen den er', 12),
 (u'kriminalit\xe4t nicht auszahlt', 12),
 (u'man sagen dass', 11),
 (u'eines menschen sollte', 11),
 (u'beitrag f\xfor die', 11),
 (u'rolle der frau', 11),
 (u'die frage ob', 11),
 (u'der feminismus hat', 11),
 (u'kann man sagen', 10),
 (u'menschen sollte dem', 10),

(u'in der lage', 10),
 (u'die wirkliche welt', 10),
 (u'den meisten f\xllen', 10),
 (u'ist es nicht', 10),
 (u'doch was ist', 10),
 (u'zahlt sich kriminalit\xe4t', 10),
 (u'sollte dem beitrag', 10),
 (u'den interessen der', 10),
 (u'ob sich kriminalit\xe4t', 10),
 (u'sich die kriminalit\xe4t', 9),
 (u'wenn man sich', 9),
 (u'auf die wirkliche', 9),
 (u'der feminismus den', 9),
 (u'sagen dass sich', 9),
 (u'auch wenn sie', 9),
 (u'die menschen die', 9),
 (u'bin der meinung', 8),
 (u'es sich um', 8),
 (u'feminismus hat den', 8),
 (u'ob es sich', 8),
 (u'auf der welt', 8),
 (u'erwischt zu werden', 8),
 (u'sich die frage', 8),
 (u'ich bin der', 8),
 (u'der ganzen welt', 8),
 (u'auf der anderen', 8),
 (u'der reiche mann', 7),
 (u'dass kriminalit\xe4t sich', 7),
 (u'hat der feminismus', 7),
 (u'die rolle der', 7),
 (u'auf keinen fall', 7),
 (u'auf jeden fall', 7),
 (u'nicht so viel', 7),
 (u'frage ob sich', 7),
 (u'zu tun haben', 7),
 (u'gibt es auch', 7),
 (u'dass sich die', 7),
 (u'nicht mehr so', 7),
 (u'sich um die', 7),
 (u'der heutigen gesellschaft', 7),
 (u'feminismus den frauen', 7),
 (u'dass es sich', 7),
 (u'dass die frau', 7),
 (u'der anderen seite', 7),
 (u'von geringem wert', 7),

(u'es gibt viele', 7),
 (u'es sich f\xfor', 7),
 (u'das was sie', 6),
 (u'dass der feminismus', 6),
 (u'die gesellschaft und', 6),
 (u'zum beispiel die', 6),
 (u'nun die frage', 6),
 (u'und so weiter', 6),
 (u'erst in den', 6),
 (u'nicht auf die', 6),
 (u'hat den interessen', 6),
 (u'der ansicht dass', 6),
 (u'ein gro\xdfes problem', 6),
 (u'frauen die sich', 6),
 (u'meinung nach ist', 6),
 (u'es sich nicht', 6),
 (u'kriminell zu werden', 6),
 (u'auch in der', 6),
 (u'frau in der', 6),
 (u'was ist mit', 6),
 (u'in den k\xpfen', 6),
 (u'in der schule', 6),
 (u'gegen das gesetz', 6),
 (u'f\xfor sich selbst', 6),
 (u'dass sie sich', 6),
 (u'die meisten universit\xe4tsabschl\xfcsse', 6),
 (u'nicht mehr nur', 5),
 (u'gibt es noch', 5),
 (u'1 die finanzielle', 5),
 (u'sich um den', 5),
 (u'eines menschen dem', 5),
 (u'die frage ist', 5),
 (u'dass es nicht', 5),
 (u'auch heute noch', 5),
 (u'in form von', 5),
 (u'die gesellschaft leisten', 5),
 (u'in den medien', 5),
 (u'mehr f\xfor die', 5),
 (u'ich finde es', 5),
 (u'das ist auch', 5),
 (u'zu diesem thema', 5),
 (u'studenten nicht auf', 5),
 (u'zahlt sich die', 5),
 (u'gro\xdf und ganzen', 5),

(u'der meinung das', 5),
(u'es in der', 5),
(u'nicht erwischt wird', 5),
(u'voll und ganz', 5),
(u'es aber auch', 5),
(u'in deutschland und', 5),
(u'hier in deutschland', 5),
(u'eine gro\xdfte rolle', 5),
(u'in der politik', 5),
(u'die interessen der', 5),
(u'sich f\xfcr die', 5),
(u'noch nicht einmal', 5),
(u'auf der ganzen', 5),
(u'nicht nur die', 5),
(u'es stellt sich', 5),
(u'menschen die sich', 5),
(u'dass die frauen', 5),
(u'im gro\xdfen und', 5),
(u'da sie ja', 5),
(u'und vieles mehr', 5),
(u'nicht praxisorientiert und', 5),
(u'doch es gibt', 5),
(u'der heutigen zeit', 5),
(u'in der praxis', 5),
(u'zu sagen dass', 5),
(u'heutzutage nicht mehr', 5),
(u'der fall ist', 5),
(u'wenn man die', 5),
(u'in der welt', 5),
(u'kriminalit\xe4t sich nicht', 5),
(u'in ihrem beruf', 5),
(u'art und weise', 5),
(u'in der familie', 5),
(u'zur rechenschaft gezogen', 5),
(u'es ist ein', 5),
(u'sich kriminalit\xe4t auszahlt', 5),
(u'die f\xfcr die', 5),
(u'muss man sich', 5),
(u'es gibt auch', 5),
(u'kann man also', 5),
(u'kriminalit\xe4t ist ein', 5),
(u'handelt es sich', 5),
(u'fr\xfcher oder sp\xe4ter', 5),
(u'f\xfcr die frauen', 5),
(u'lohnt es sich', 5),

Liste 6. L1_n4_mostcommon

```
[('zahlt sich nicht aus', 27),
 ('kriminalit\xe4t zahlt sich nicht',
 24),
 ('sie f\xfc die gesellschaft', 19),
 ('f\xfc die gesellschaft geleistet',
 18),
 ('die gesellschaft geleistet hat',
 16),
 ('finanzielle entlohnung eines
 menschen', 16),
 ('die finanzielle entlohnung eines',
 16),
 ('den er sie f\xfc', 15),
 ('er sie f\xfc die', 15),
 ('mehr geschadet als gen\xfcftzt',
 14),
 ('frauen mehr geschadet als', 13),
 ('in den letzten jahren', 13),
 ('der frauen mehr geschadet', 12),
 ('beitrag entsprechen den er', 12),
 ('entlohnung eines menschen sollte',
 11),
 ('dem beitrag entsprechen den', 11),
 ('entsprechen den er sie', 11),
 ('beitrag f\xfc die gesellschaft',
 11),
 ('den interessen der frauen', 10),
 ('in den meisten f\xe4llen', 10),
 ('eines menschen sollte dem', 10),
 ('menschen sollte dem beitrag', 10),
 ('sollte dem beitrag entsprechen',
 10),
 ('sich kriminalit\xe4t nicht
 auszahlt', 9),
 ('interessen der frauen mehr', 9),
 ('kann man sagen dass', 8),
 ('dass sich kriminalit\xe4t nicht',
 8),
 ('der feminismus hat den', 8),
 ('auf die wirkliche welt', 8),
 ('ich bin der meinung', 7),
 ('die rolle der frau', 7),
 ('in der heutigen gesellschaft', 7),
```

```
('auf der anderen seite', 7),
 ('nicht auf die wirkliche', 6),
 ('meiner meinung nach ist', 6),
 ('der feminismus den frauen', 6),
 ('hat den interessen der', 6),
 ('feminismus hat den interessen', 5),
 ('zahlt sich die kriminalit\xe4t',
 5),
 ('man sagen dass sich', 5),
 ('in der heutigen zeit', 5),
 ('f\xfc die gesellschaft und', 5),
 ('im gro\xdfen und ganzen', 5),
 ('frage ob sich kriminalit\xe4t', 5),
 ('f\xfc die gesellschaft leisten',
 5),
 ('l die finanzielle entlohnung', 5),
 ('auf der ganzen welt', 5),
 ('studenten nicht auf die', 5),
```


Liste 7. L1_n3_PMI

[(u'voll und ganz', 8.535230607705435),	(u'handelt es sich',	(u'auf keinen fall',
(u'fr\xfcher oder sp\xeter',	6.000023219442635),	4.969249289914652),
8.124488442496427),	(u'rolle der frau',	(u'eine gro\xdfen rolle',
(u'art und weise',	5.969060993838668),	4.968608469642803),
7.6349402171777205),	(u'erwischt zu werden',	(u'auf die wirkliche',
(u'geschadet als gen\xftzt',	5.953932112242368),	4.902109987077023),
7.572419860196386),	(u'frage ob sich', 5.681814556331699),	(u'er sie f\xfcr', 4.888220874916356),
(u'l die finanzielle',	(u'gesellschaft geleistet hat',	(u'kriminalit\xet zahlt sich',
7.420869962069186),	5.639581792708426),	4.883233333405317),
(u'entsprechen den er',	(u'feminismus den frauen',	(u'ob sich kriminalit\xet',
7.3738174177982945),	5.593061777073533),	4.828651576492097),
(u'finanzielle entlohnung eines',	(u'frauen mehr geschadet',	(u'heutzutage nicht mehr',
7.367625447550373),	5.532437155257098),	4.769978174806924),
(u'gro\xdfen und ganzen',	(u'sollte dem beitrag',	(u'doch was ist', 4.726629962099786),
7.355355354958559),	5.521798757052043),	(u'den letzten jahren',
(u'beitrag entsprechen den',	(u'kriminell zu werden',	4.725046987536536),
7.117683703081439),	5.465579344328436),	(u'den meisten f\xellen',
(u'meiner meinung nach',	(u'dem beitrag entsprechen',	4.725046987536535),
7.090284728893325),	5.42778051799815),	(u'hat den interessen',
(u'studenten nicht auf',	(u'zahlt sich kriminalit\xet',	4.682048102300222),
6.994785566758285),	5.390462754990955),	(u'ein gro\xdfes problem',
(u'entlohnung eines menschen',	(u'menschen sollte dem',	4.664801499576125),
6.852885924463246),	5.322724231122713),	(u'den ganzen tag',
(u'zur rechenschaft gezogen',	(u'sagen dass sich',	4.656054116049584),
6.834820917065607),	5.307575095185188),	(u'hier in deutschland',
(u'interessen der frauen',	(u'mehr geschadet als',	4.654485341362089),
6.702649143957124),	5.30452032519023),	(u'erst in den', 4.632777053293439),
(u'meinung nach ist',	(u'gibt es noch', 5.290656095015926),	(u'zahlt sich die',
6.4604294455821005),	(u'lohnt es sich', 5.274086216059699),	4.632777053293439),
(u'eines menschen dem',	(u'im gro\xdfen und',	(u'auf der anderen',
6.453806904821607),	5.210994267179407),	4.617272866757474),
(u'eines menschen sollte',	(u'kann man sagen',	(u'noch nicht einmal',
6.453806904821607),	5.126390548178308),	4.608084440703068),
(u'gegen das gesetz',	(u'ich bin der', 5.104705610948759),	(u'den er sie', 4.599883844582529),
6.232645514711389),	(u'auf jeden fall',	(u'f\xfcr die gesellschaft',
(u'bin der meinung',	5.102780682539175),	4.546834623492714),
6.214945937611987),	(u'ich denke dass',	(u'beitrag f\xfcr die',
(u'feminismus hat den',	5.088957253980619),	4.518366702115695),
6.106082791402959),	(u'was ist mit', 5.081638245965365),	(u'auch heute noch',
(u'zum beispiel die',	(u'da sie ja', 5.070983908457175),	4.445659324235991),
6.092193679242292),	(u'von geringem wert',	(u'ob es sich', 4.441878601396085),
(u'zahlt sich nicht',	5.057660247258705),	(u'kann man also',
6.0159897478759214),	(u'ich finde es', 5.040167089811188),	4.4332433676183625),
	(u'kriminalit\xet nicht auszahlt',	(u'dass kriminalit\xet sich',
	5.01840811177357),	4.3516410851820995),

(u'auf der welt', 4.349008880162794),
 (u'nun die frage',
 4.3355255298255075),
 (u'in den medien', 4.330774927066909),
 (u'in der lage', 4.330774927066909),
 (u'in deutschland und',
 4.330774927066909),
 (u'in den letzten',
 4.330774927066909),
 (u'in den k\xf6pfen',
 4.330774927066909),
 (u'gibt es auch', 4.240833970517248),
 (u'zu tun haben', 4.2326681443858),
 (u'auf der ganzen',
 4.227311945185274),
 (u'es gibt viele', 4.215628328476789),
 (u'in der heutigen',
 4.187674083426236),
 (u'sich kriminalit\xe4t auszahlt',
 4.150825850626585),
 (u'in form von', 4.1484533702729545),
 (u'in der politik',
 4.1484533702729545),
 (u'in ihrem beruf',
 4.1484533702729545),
 (u'den interessen der',
 4.1372603226344165),
 (u'in unserer gesellschaft',
 4.136618912625952),
 (u'nicht mehr nur', 4.04392258263049),
 (u'nicht praxisorientiert und',
 4.04392258263049),
 (u'es gibt auch', 4.033306771682834),
 (u'in den meisten',
 3.9943026904456964),
 (u'in der praxis',
 3.9943026904456964),
 (u'nicht mehr so', 3.974929711143538),
 (u'sich nicht aus',
 3.965422627295222),
 (u'sich kriminalit\xe4t nicht',
 3.9276822993123757),
 (u'in der schule',
 3.9253098189587448),
 (u'zu diesem thema',
 3.8961959077645867),
 (u'der meinung das',
 3.864034911458193),
 (u'der heutigen zeit',
 3.864034911458193),
 (u'der fall ist', 3.864034911458193),
 (u'der ganzen welt',
 3.864034911458193),
 (u'der anderen seite',
 3.864034911458193),
 (u'dass sich kriminalit\xe4t',
 3.8301770632110532),
 (u'kriminalit\xe4t ist ein',
 3.829281815564628),
 (u'wenn man die', 3.829138312658448),
 (u'frauen die sich',
 3.804780572221803),
 (u'der meinung dass',
 3.8034102896417585),
 (u'f\xfcr sich selbst',
 3.773644735259232),
 (u'der feminismus den',
 3.758674395800367),
 (u'muss man sich', 3.743778463707291),
 (u'der reiche mann',
 3.730503518833671),
 (u'wenn man sich',
 3.7237777970006216),
 (u'der ansicht dass',
 3.709884231630935),
 (u'der frauen mehr',
 3.640891360143984),
 (u'wie zum beispiel',
 3.6395549578365403),
 (u'der feminismus hat',
 3.5538799831543537),
 (u'man sagen dass',
 3.552790928049438),
 (u'und so weiter',
 3.5382686155093883),
 (u'die gesellschaft geleistet',
 3.4984841696054927),
 (u'die wirkliche welt',
 3.4984841696054927),
 (u'die meisten
 universit\xe4tsabschl\xfcssen',
 3.4984841696054927),
 (u'nicht erwischt wird',
 3.437786779060174),
 (u'der heutigen gesellschaft',
 3.412049787715136),
 (u'hat der feminismus',
 3.383946453215792),
 (u'und vieles mehr',
 3.355947058715434),
 (u'nicht nur die',
 3.3507754020705445),
 (u'die gesellschaft leisten',
 3.3161626128115382),
 (u'die finanzielle entlohnung',
 3.3092421699669643),
 (u'nicht so viel', 3.281782530583593),
 (u'es sich f\xfcr',
 3.271166719635937),
 (u'sich um den', 3.2345351187524303),
 (u'in der familie', 3.2321626383988),
 (u'menschen die sich',
 3.218590076852506),
 (u'die frage ob', 3.1883292413016533),
 (u'auch wenn sie',
 3.1616438122365187),
 (u'dass sie sich', 3.161398473668066),
 (u'in der gesellschaft',
 3.1196846549721093),
 (u'in der welt', 3.1069994954447937),
 (u'das ist auch', 3.053269807105527),
 (u'die rolle der',
 3.0464990458624355),
 (u'es sich um', 3.036973332135142),
 (u'die frage ist', 3.028480540359757),
 (u'dass die frau',
 3.0118667396971017),
 (u'f\xfcr die frauen',
 2.9905493180363476),
 (u'dass die frauen',
 2.979076916874111),
 (u'das was sie', 2.9732270994319907),
 (u'es stellt sich',
 2.934694483014724),
 (u'kriminalit\xe4t sich nicht',

```
2.912991083690473),
(u'doch es gibt', 2.9062971204276744),
(u'ist es nicht', 2.8963676791485486),
(u'dass sich die', 2.878335347072579),
(u'sich die kriminalit\xe4t',
2.8167999180524514),
(u'sich um die', 2.8133216536761263),
(u'dass es sich', 2.779244444428348),
(u'dass der feminismus',
2.745238076443153),
(u'die gesellschaft und',
2.725294281372011),
(u'dass es nicht',
2.6994920546549492),
(u'frau in der', 2.600924474437401),
(u'sie f\xfcr die',
2.576651630004505),
(u'es ist ein', 2.5669697028894065),
(u'sich die frage',
2.541387938192485),
(u'es aber auch', 2.457770410924415),
(u'zu sagen dass', 2.408118852334754),
(u'die interessen der',
2.2175503241434282),
(u'mehr f\xfcr die',
1.9773705856766508),
(u'nicht auf die',
1.9236590464303989),
(u'es sich nicht',
1.8330005278092063),
(u'die menschen die',
1.7836857415135658),
(u'auch in der', 1.2839419132077392),
(u'sich f\xfcr die',
0.9453730460905249),
(u'es in der', 0.689267803860627),
(u'die f\xfcr die',
0.11070980827547798)]
```

Liste 8. L1_n4_PMI

[('studenten nicht auf die',
7.845620854910474),
(u'entsprechen den er sie',
7.626815705194863),
(u'finanzielle entlohnung eines
menschen', 7.428321289155307),
(u'l die finanzielle entlohnung',
7.424407389834171),
(u'entlohnung eines menschen sollte',
7.201263838519961),
(u'interessen der frauen mehr',
7.201263838519961),
(u'meiner meinung nach ist',
7.121221130846425),
(u'beitrag entsprechen den er',
7.121221130846425),
(u'zahlt sich nicht aus',
6.898077579532215),
(u'eines menschen sollte dem',
6.457344332586593),
(u'feminismus hat den interessen',
6.412806478155691),
(u'zahlt sich die kriminalit\xe4t',
6.310290914630096),
(u'frage ob sich kriminalit\xe4t',
5.9366664123775905),
(u'sollte dem beitrag entsprechen',
5.83071783436821),
(u'im gro\xdfen und ganzen',
5.551003931565606),
(u'frauen mehr geschadet als',
5.522491232684796),
(u'dem beitrag entsprechen den',
5.413299440260458),
(u'mehr geschadet als gen\xfcgt',
5.368682374771651),
(u'kriminalit\xe4t zahlt sich nicht',
5.363694833260612),
(u'beitrag f\xfc die gesellschaft',
5.3447291337491585),
(u'menschen sollte dem beitrag',
5.326261658887699),
(u'er sie f\xfc die',

5.214531694944393),
(u'auf der anderen seite',
5.10631811030416),
(u'ich bin der meinung',
5.092494681745605),
(u'kann man sagen dass',
5.07383850929225),
(u'auf die wirkliche welt',
4.883174558989951),
(u'f\xfc die gesellschaft geleistet',
4.81863603785238),
(u'f\xfc die gesellschaft leisten',
4.818636037852379),
(u'hat den interessen der',
4.780895709869532),
(u'den er sie f\xfc',
4.728584415301521),
(u'f\xfc die gesellschaft und',
4.636314481058425),
(u'auf der ganzen welt',
4.636314481058425),
(u'in der heutigen zeit',
4.334312354831895),
(u'in den meisten f\xe4llen',
4.334312354831895),
(u'in der heutigen gesellschaft',
4.334312354831895),
(u'in den letzten jahren',
4.334312354831895),
(u'den interessen der frauen',
4.258580786055785),
(u'dass sich kriminalit\xe4t nicht',
4.24754784875472),
(u'sich kriminalit\xe4t nicht
auszahlt', 4.049002762733744),
(u'man sagen dass sich',
4.00463362259908),
(u'der feminismus hat den',
3.867572339223179),
(u'nicht auf die wirkliche',
3.8243164590812655),
(u'der frauen mehr geschadet',
3.713421659395921),
(u'der feminismus den frauen',
3.713421659395921),

(u'die gesellschaft geleistet hat',
3.502021597370478),
(u'die finanzielle entlohnung eines',
3.502021597370478),
(u'sie f\xfc die gesellschaft',
3.302930603080986),
(u'die rolle der frau',
3.0500364736274213)]

Tabelle 1. L2_n3_mostcommon

Chunk	Frequenz	WV ¹	WVM ²	Merkmal	Kriterium
meiner meinung nach	132	1	0	gesprächsspezifische Formel, meinungsäußernd	pragmatisch
in der gesellschaft	115	0	1	Präpositionalphrase, lokal, nicht zielgerichtet, übertragener Gebrauch	strukturell
an der universität	96	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
an der uni	54	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
es gibt viele	49	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
in der welt	45	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
ich glaube dass	45	0	1	gesprächsspezifische Formel, meinungsäußernd	pragmatisch
männer und frauen	43	0	1	Paarformel, thematisch bedingt	strukturell
es gibt auch	41	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
auf jeden fall	35	1	0	Präpositionalphrase	strukturell
frauen und männer	34	0	1	Paarformel, fehlerhaft?	strukturell
aber es gibt	34	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
in der arbeitswelt	33	0	1	Präpositionalphrase, lokal, nicht zielgerichtet, übertragener Gebrauch	strukturell
in der vergangenheit	32	0	1	Präpositionalphrase, temporal	strukturell
aber es ist	31	0	1	gesprächsspezifische Formel, evaluativ?	pragmatisch
und es gibt	31	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
in der schule	30	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
und so weiter	30	1	0	gesprächsspezifische Formel, sinnentleert, fehlerhaft?	pragmatisch
in bezug auf	29	1	0	Präpositionalphrase	strukturelles Phrasem
zu hause bleiben	28	1	0	Kollokation	semantisch
in der zukunft	27	0	1	Präpositionalphrase, temporal	strukturell
nach dem studium	27	0	1	Präpositionalphrase, temporal	strukturell
wie zum beispiel	25	1	0	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
in unserer gesellschaft	25	0	1	Präpositionalphrase, lokal, nicht zielgerichtet, übertragener Gebrauch	pragmatisch

¹ Wortverbindung² Wortverbindungs muster

Chunk	Frequenz	WV ¹	WVM ²	Merkmal	Kriterium
in diesem fall	24	1	0	Präpositionalphrase	strukturell
in der universität	23	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
bei der arbeit	23	0	1	Präpositionalphrase, temporal, Gleichzeitigkeit, Zeitdauer	strukturell
die mehrheit der	21	0	1	Nominalphrase, Genitivattribut	strukturell
in den usa	21	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
im vergleich zu	21	1	0	Präpositionalphrase, strukturelles Phrasem	strukturell
in der praxis	20	1	0	Präpositionalphrase, übertragener Gebrauch	strukturell
in diesem aufsatz	19	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
am ende des	19	0	1	Nominalphrase, Genitivattribut	strukturell
zum beispiel wenn	19	1	0	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
ein großes problem	18	0	1	Nominalphrase, Adjektivattribut	strukturell
ich denke dass	18	0	1	gesprächsspezifische Formel, meinungsäußernd	pragmatisch
die frage ist	18	0	1	gesprächsspezifische Formel, thematisierend	pragmatisch
zum beispiel in	18	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
die meisten studenten	18	0	1	Nominalphrase, Adjektivattribut, quantifizierend	strukturell
es gibt aber	18	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
während des studiums	18	0	1	Präpositionalphrase, temporal, Gleichzeitigkeit	strukturell
es ist aber	18	0	1	gesprächsspezifische Formel, evaluativ?	pragmatisch
im bezug auf	17	1	0	fehlerhaft	strukturell
meine meinung nach	16	1	0	fehlerhaft	strukturell
man könnte sagen	16	0	1	gesprächsspezifische Formel, relativierend	pragmatisch
deshalb ist es	16	0	1	gesprächsspezifische Formel, argumentativ	pragmatisch
zum beispiel die	16	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
die mehrheit von	15	0	1	Nominalphrase, Ersatzform des Genitivs, fehlerhaft?	strukturell
die rolle der	15	0	1	Nominalphrase, Genitivattribut	strukturell
im vergleich mit	15	1	0	Präpositionalphrase, strukturelles Phrasem	strukturell
das leben der	15	0	1	Nominalphrase, Genitivattribut	strukturell
die gleiche rechte	14	0	1	fehlerhaft	strukturell

Chunk	Frequenz	WV ¹	WVM ²	Merkmal	Kriterium
es gibt noch	14	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
die meisten universitätsabschlüsse	14	0	1	fehlerhaft	strukturell
es ist klar	14	0	1	gesprächsspezifische Formel, evaluativ	pragmatisch
die frage ob	14	0	1	gesprächsspezifische Formel, thematisierend	pragmatisch
auf der straße	14	0	1	Präpositionalphrase, lokal, mit Berührung, nicht zielgerichtet	strukturell
an der wirtschaftsuniversität	14	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
wegen des feminismus	13	0	1	Präpositionalphrase, kausal	strukturell
nach dem abschluss	13	0	1	Präpositionalphrase, temporal	strukturell
so viel geld	13	0	1	Nominalphrase, unbestimmtes Zahladjektiv	strukturell
das problem ist	13	0	1	gesprächsspezifische Formel, evaluativ	pragmatisch
die neuen studiengänge	13	0	1	Nominalphrase, Adjektivattribut	strukturell
von geringen wert	13	0	1	fehlerhaft	strukturell
die tatsache dass	12	0	1	gesprächsspezifische Formel, thematisierend	pragmatisch
in der arbeit	12	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
es kann sein	12	0	1	gesprächsspezifische Formel, relativierend	pragmatisch
für die arbeitswelt	12	0	1	Präpositionalphrase, modal, Komparativ (im Hinblick auf)	strukturell
ab und zu	12	1	0	Paarformel	strukturell
an einer universität	12	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
heutzutage gibt es	12	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
das wort feminismus	12	0	1	Nominalphrase	strukturell
eine gute arbeit	12	0	1	Nominalphrase, Adjektivattribut	strukturell
das neue system	12	0	1	Nominalphrase, Adjektivattribut	strukturell
in der uni	12	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
im laufe der	12	0	1	Präpositionalphrase, temporal	strukturell

Tabelle 2. L2_n4_mostcommon

Chunk	Frequenz	WV	WVM	Merkmal	Kriterium
auf der anderen seite	47	1	0	Präpositionalphrase, strukturelles Phrasem	strukturell
in der wirklichen welt	42	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
die interessen der frauen	24	0	1	Nominalphrase, Genitivattribut	strukturell
ich bin der meinung	17	0	1	gesprächsspezifische Formel, meinungsäußernd	pragmatisch
man könnte sagen dass	14	0	1	gesprächsspezifische Formel, relativierend	pragmatisch
in den letzten jahren	14	0	1	Präpositionalphrase, temporal	strukturell
in der heutigen gesellschaft	13	0	1	Präpositionalphrase, lokal, nicht zielgerichtet, übertragener Gebrauch	strukturell
zum beispiel wenn man	13	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
es ist klar dass	13	0	1	gesprächsspezifische Formel, evaluativ	pragmatisch
in die wirkliche welt	13	0	1	Präpositionalphrase, lokal, zielgerichtet	strukturell
zwischen männern und frauen	12	0	1	Präpositionalphrase, übertragener Gebrauch	strukturell
auf der einen seite	11	1	0	Präpositionalphrase, strukturelles Phrasem	strukturell
zwischen männer und frauen	11	0	1	fehlerhaft	strukturell
es kann sein dass	10	0	1	gesprächsspezifische Formel, relativierend	pragmatisch
die interesse der frauen	10	0	1	fehlerhaft	strukturell
das leben der frauen	9	0	1	Nominalphrase, Genitivattribut	strukturell
auf der andere seite	9	1	0	fehlerhaft	strukturell
im großen und ganzen	9	1	0	Paarformel	strukturell
die rechte der frauen	9	0	1	Nominalphrase, Genitivattribut	strukturell
es ist nicht nur	8	0	1	strukturelles Phrasem (nicht nur... sondern?)	strukturell
um geld zu verdienen	8	0	1	Finalsatz "um...zu", enthält Kollokation Geld verdienen	strukturell
zum schluss glaube ich	7	0	1	fehlerhaft?	pragmatisch
ich würde sagen dass	7	0	1	gesprächsspezifische Formel, meinungsäußernd	pragmatisch
die rolle der frauen	7	0	1	Nominalphrase, Genitivattribut	strukturell
in deutschland gibt es	7	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
in den meisten fällen	7	1	0	Präpositionalphrase	strukturell

Chunk	Frequenz	WV	WVM	Merkmal	Kriterium
das bedeutet nicht dass	7	0	1	gesprächsspezifische Formel, relativierend	pragmatisch
in der wirkliche welt	7	0	1	fehlerhaft	strukturell
es kommt darauf an	6	0	1	gesprächsspezifische Formel, relativierend	pragmatisch
wenn man z. b.	6	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
es ist möglich dass	6	0	1	gesprächsspezifische Formel, relativierend	pragmatisch
zum schluss möchte ich	6	0	1	strukturelles Phrasem	strukturell
aber es gibt auch	6	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
wenn man zum beispiel	6	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
aber ich denke dass	6	0	1	gesprächsspezifische Formel, meinungsäußernd	pragmatisch
es gibt viele gründe	5	0	1	gesprächsspezifische Formel, argumentativ	pragmatisch
zum beispiel gibt es	5	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
es ist wahr dass	5	0	1	gesprächsspezifische Formel, evaluativ	pragmatisch
man darf nicht vergessen	5	0	1	gesprächsspezifische Formel, thematisierend	pragmatisch
z. b. wenn man	5	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
die situation der frauen	5	0	1	Nominalphrase, Genitivattribut	strukturell

Tabelle 3. L2_n3_PMI

Chunk	PMI	WV	WVM	Merkmal	Kriterium
simone de beauvoir	10,1	1	0	Eigennamen	semantisch
vor- und nachteile	9,2	1	0	Paarformel	strukturell
u. s. w.	8,9				
d. h. dass	8,7	0	1	gesprächsspezifisch, erklärend	pragmatisch
art und weise	8,3	1	0	Paarformel	strukturell
vorteile und nachteile	7,5	1	0	Paarformel	strukturell
z. b. wenn	7,4				
typen von kriminalität	7,4	0	1	Nominalphrase, Ersatzform des Genitivs, Nullartikel	strukturell
ins gefängnis gehen	7,3	0	1	Präpositionalphrase, Kollokation, lokal, zielgerichtet	strukturell
seit vielen jahren	7,3	0	1	Präpositionalphrase, temporal	strukturell
zwischen den geschlechtern	7,1			fehlerhaft	strukturell
ihr eigenes geld	7	0	1	Nominalphrase, Adjektivattribut, verstärkt Possessivartikel	strukturell
meiner meinung nach	6,9				
egal ob man	6,8	0	1	Subjunktion ob...oder (ob)?, strukturelles Phrasem	strukturell
am wichtigsten ist	6,8	0	1	gesprächsspezifisch, evaluativ	pragmatisch
ab und zu	6,8				
alles in allem	6,5	1	0	Paarformel	strukturell
am ende des	6,4				
während des studiums	6,1				
bei einer firma	6	0	1	Präpositionalphrase, lokal, volkswirtschaftlicher Bereich	strukturell
am anfang des	5,9	0	1	Präpositionalphrase, temporal	strukturell
vor 50 jahren	5,9	0	1	Präpositionalphrase, temporal	strukturell
nach dem abschluss	5,9				
zum ersten mal	5,8	0	1	Präpositionalphrase, temporal	strukturell
an der wirtschaftsuniversität	5,7				
zum beispiel wenn	5,7				

Chunk	PMI	WV	WVM	Merkmal	Kriterium
aus diesem grund	5,7	1	0	gesprächsspezifisch, argumentativ	pragmatisch
im rahmen der	5,6	0	1	Präpositionalphrase, Nominalphrase, Genitivattribut, übertragener Gebrauch	strukturell
im laufe des	5,6				
immer mehr studenten	5,6	0	1	Zahlwort, Nominalphrase	strukturell
im vergleich mit	5,6				
dies bedeutet dass	5,6	0	1	gesprächsspezifisch, erklärend	pragmatisch
nach dem studium	5,5				
im laufe der	5,5				
im vergleich zu	5,5				
immer mehr leute	5,5	0	1	Zahlwort, Nominalphrase	strukturell
männer und frauen	5,4				
mehr oder weniger	5,4	1	0	Paarformel	strukturell
an der uni	5,3				
auf jeden fall	5,3				
an einer universität	5,2				
an den universitäten	5,2				
auf keinen fall	5,2				
immer mehr frauen	5,1	0	1	Zahlwort, Nominalphrase	strukturell
wie schon erwähnt	5,1	0	1	gesprächsspezifisch, textstrukturierend	pragmatisch
nach ihrem studium	5,1				
auf der straße	5				
meine meinung nach	5				
auf jedem fall	5			fehlerhaft	strukturell
ich meine dass	5			fehlerhaft?	pragmatisch
ich hoffe dass	5	0	1	gesprächsspezifisch, meinungsäußernd	pragmatisch
den ganzen tag	5	0	1	Adverbialbestimmung, Nominalphrase, temporal	strukturell
an der universitäten	5			fehlerhaft	strukturell
ich denke dass	5				

Chunk	PMI	WV	WVM	Merkmal	Kriterium
auf dem arbeitsmarkt	5	0	1	Präpositionalphrase, lokal, nicht zielgerichtet, übertragener Gebrauch	strukturell
von geringerem wert	5	0	1	Präpositionalphrase, komparativ, Dativ	strukturell
ich glaube dass	5				
ich finde es	4,9	0	1	gesprächsspezifisch, evaluativ?	strukturell
ein gutes beispiel	4,8	0	1	Nominalphrase, Adjektivattribut	strukturell
ein halbes jahr	4,9	0	1	Nominalphrase, Adjektivattribut	strukturell
ich finde dass	4,8	0	1	gesprächsspezifisch, meinungsäußernd	pragmatisch
eine gute idee	4,8	0	1	Nominalphrase, Adjektivattribut	strukturell
mit dem geld	4,8	0	1	Präpositionalphrase, modal	strukturell
mit den kindern	4,7	0	1	Präpositionalphrase, partitiv, Zugehörigkeit	strukturell
ich würde sagen	4,7				
ein besseres leben	4,7	0	1	Nominalphrase, Adjektivattribut, komparativ	strukturell

Tabelle 4. L2_n4_PMI

Chunk	PMI	WV	WVM	Merkmal	Kriterium
z. b. wenn man	7,4				
zwischen männern und frauen	7				
überall in der welt	6,8	0	1	Adverb, Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
zum schluss glaube ich	5,8				
an der wirtschaftsuniversität kopenhagen	5,8	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
könnte gesagt werden dass	5,7	0	1	gesprächsspezifische Formel, relativierend	pragmatisch
im großen und ganzen	5,7				
an der universität studieren	5,6	0	1	Präpositionalphrase, Kollokation, lokal, nicht zielgerichtet	strukturell, semantisch
auf der anderen seite	5,3				
auf der andere seite	5,3				
auf der einen seite	5,2				
wenn man z. b.	5				
ich bin der meinung	5				
ich würde sagen dass	4,9				
für mich ist es	4,8	0	1	gesprächsspezifische Formel, evaluativ?	pragmatisch
das leben der frauen	4,7				
das bedeutet nicht dass	4,7				
in deutschland gibt es	4,2				
in den meisten fällen	4,2				
in den letzten jahren	4,2				
man könnte sagen dass	4,2				
es kann sein dass	4,2				
es gibt viele gründe	4,2				
es ist möglich dass	4,2				
in der heutigen gesellschaft	4,2				
es ist klar dass	4,1				

Chunk	PMI	WV	WVM	Merkmal	Kriterium
es kommt darauf an	4				
frauen in der politik	4	0	1	Nominalphrase	strukturell
es ist wahr dass	4				
wenn man zum beispiel	4				
aber ich denke dass	3,9				
in der wirklichen welt	3,8				
in der wirkliche welt	3,6				
es ist nicht nur	3,4				
die situation der frauen	3,2				
die rechte der frauen	3				
die rolle der frauen	2,9				
in die wirkliche welt	1,1				
die interessen der frauen	0,9				

Tabelle 5. L1_n3_mostcommon

Chunk	Frequenz	WV	WVM	Merkmal	Kriterium
meiner meinung nach	36	1	0	gesprächsspezifische Formel, meinungsäußernd	pragmatisch
ich denke dass	14	0	1	gesprächsspezifische Formel, meinungsäußernd	pragmatisch
in unserer gesellschaft	14	0	1	Präpositionalphrase, lokal, nicht zielgerichtet, übertragener Gebrauch	strukturell
in der gesellschaft	14	0	1	Präpositionalphrase, lokal, nicht zielgerichtet, übertragener Gebrauch	strukturell
wie zum beispiel	12	1	0	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
die frage ob	11	0	1	gesprächsspezifische Formel, thematisierend	pragmatisch
in der lage	10		1	Präpositionalphrase, übertragener Gebrauch, fähig sein	strukturell
auf der welt	8	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
der reiche mann	7	0	1	Nominalphrase, Adjektivattribut	strukturell
auf keinen fall	7	1	0	Präpositionalphrase	strukturell
auf jeden fall	7	1	0	Präpositionalphrase	strukturell
es gibt viele	7	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
und so weiter	6	1	0	gesprächsspezifische Formel, sinnentleert	pragmatisch
ein großes problem	6	0	1	Nominalphrase, Adjektivattribut	strukturell
in den köpfen	6	1	0	Präpositionalphrase, lokal, nicht zielgerichtet, übertragener Gebrauch	strukturell
in der schule	6	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
gegen das gesetz	6	0	1	Präpositionalphrase, modal	strukturell
in form von	5	1	0	Präpositionalphrase	strukturell
in den medien	5	0	1	Präpositionalphrase, lokal, nicht zielgerichtet, übertragener Gebrauch	strukturell
ich finde es	5	0	1	gesprächsspezifische Formel, meinungsäußernd	pragmatisch
zu diesem thema	5	0	1	Präpositionalphrase, bezüglich	strukturell
voll und ganz	5	1	0	Paarformel	strukturell
hier in deutschland	5	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
eine große rolle	5	0	1	Nominalphrase, Adjektivattribut	strukturell
in der politik	5	0	1	Präpositionalphrase, lokal, nicht zielgerichtet, übertragener Gebrauch	strukturell
und vieles mehr	5	1	0	gesprächsspezifische Formel, sinnentleert	pragmatisch

Chunk	Frequenz	WV	WVM	Merkmal	Kriterium
doch es gibt	5	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
in der praxis	5	1	0	Präpositionalphrase, übertragener Gebrauch	strukturell
in der welt	5	0	1	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell
in ihrem beruf	5	0	1	Präpositionalphrase, lokal, nicht zielgerichtet, übertragener Gebrauch	strukturell
art und weise	5	1	0	Paarformel	strukturell
in der familie	5	0	1	Präpositionalphrase, lokal, nicht zielgerichtet, übertragener Gebrauch	strukturell
früher oder später	5	1	0	Paarformel	strukturell

Tabelle 6. L1_n4_mostcommon

Chunk	Frequenz	WV	WVM	Merkmal	Kriterium
in den letzten jahren	13	0	1	Präpositionalphrase, temporal	strukturell
in den meisten fällen	10	1	0	Präpositionalphrase	strukturell
ich bin der meinung	7	0	1	gesprächsspezifische Formel, meinungsausßernd	pragmatisch
die rolle der frau	7	0	1	Nominalphrase, Genitivattribut	strukturell
in der heutigen gesellschaft	7	0	1	Präpositionalphrase, lokal, nicht zielgerichtet, übertragener Gebrauch	strukturell
auf der anderen seite	7	1	0	Präpositionalphrase, strukturelles Phrasem	strukturell
in der heutigen zeit	5	1	0	Präpositionalphrase, temporal	strukturell
im großen und ganzen	5	1	0	Paarformel	strukturell
auf der ganzen welt	5	1	0	Präpositionalphrase, lokal, nicht zielgerichtet	strukturell

Tabelle 7. L1_n3_PMI

Chunk	PMI	WV	WVM	Merkmal	Kriterium
voll und ganz	8,5				
früher oder später	8,1				
art und weise	7,6				
meiner meinung nach	7,1				
gegen das gesetz	6,2				
auf jeden fall	5,1				
ich denke dass	5,1				
was ist mit	5,1	0	1	gesprächsspezifische Formel, thematisierend	pragmatisch
ich finde es	5				
auf keinen fall	5				
eine große rolle	5				
ein großes problem	4,7				
den ganzen tag	4,7	0	1	Adverbialbestimmung, Nominalphrase, temporal	strukturell
hier in deutschland	4,7				
auf der welt	4,3				
in den medien	4,3				
in der lage	4,3				
in den köpfen	4,3				
es gibt viele	4,2				
in form von	4,1				
in der politik	4,1				
in ihrem beruf	4,1				
in unserer gesellschaft	4,1				
es gibt auch	4	0	1	gesprächsspezifische Formel, Beispiel gebend	pragmatisch
in der praxis	4				
in der schule	3,9				

Chunk	PMI	WV	WVM	Merkmal	Kriterium
zu diesem thema	3,9				
der reiche mann	3,7				
wie zum beispiel	3,6				
und so weiter	3,5				
und vieles mehr	3,4				
in der familie	3,2				
die frage ob	3,2				
in der gesellschaft	3,1				
in der welt	3,1				
die frage ist	3	0	1	gesprächsspezifische Formel, thematisierend	pragmatisch
doch es gibt	2,9				

Tabelle 8. L1_n4_PMI

Chunk	PMI
im großen und ganzen	5,6
auf der anderen seite	5,1
ich bin der meinung	5,1
auf der ganzen welt	4,6
in der heutigen zeit	4,3
in den meisten fällen	4,3
in der heutigen gesellschaft	4,3
in den letzten jahren	4,3
die rolle der frau	3,1

(a) Wortverbindungsmuster

(1) /[Dd]ie/ . /[Mm]ehrheit/ . "der" . pos="NN" (16)

1) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-121 (tokens 182 - 195)	die Mehrheit der Frauen
2) Path: falkoEssayL2v2.4 > fkb042_2008_08_L2v2.4 (tokens 1 - 9)	die Mehrheit der Universitäten
3) Path: falkoEssayL2v2.4 > fkb011_2008_07_L2v2.4 (tokens 363 - 376)	die Mehrheit der Sachen
4) Path: FalkoEssayL2WHIGv2.0 > BNG-2010-03-025 (tokens 256 - 269)	die Mehrheit der Uni-Studenten
5) Path: FalkoEssayL2WHIGv2.0 > BNG2-2011-03-204 (tokens 568 - 581)	die Mehrheit der Bevölkerung

(2) "in" . "der" . pos="NN" (517)

1) Path: falkoEssayL2v2.4 > fu120_2006_10a_L2v2.4 (tokens 392 - 404)	in der Grundschule
2) Path: falkoEssayL2v2.4 > fkb049_2008_08_L2v2.4 (tokens 365 - 377)	in der Praxis
3) Path: falkoEssayL2v2.4 > fkb007_2006_07_L2v2.4 (tokens 789 - 801)	in der Gesellschaft
4) Path: falkoEssayL2v2.4 > hu012_2006_09_L2v2.4 (tokens 65 - 77)	in der Gesellschaft
5) Path: falkoEssayL2v2.4 > fkb023_2008_07_L2v2.4 (tokens 261 - 273)	in der Welt

(3) /[lɪ]mmer/ . "mehr" . pos="NN" (44)

1) Path: FalkoEssayL2WHIGv2.0 > BNG-2010-03-021 (tokens 25 - 37)	immer mehr Leute
2) Path: FalkoEssayL2WHIGv2.0 > BNG-2010-04-074 (tokens 389 - 401)	immer mehr Geld
3) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-112 (tokens 291 - 303)	immer mehr Feminismus-Kritik
4) Path: falkoEssayL2v2.4 > fkb009_2007_09_L2v2.4 (tokens 3 - 17)	immer mehr Menschen
5) Path: FalkoEssayL2WHIGv2.0 > BNG-2010-03-061 (tokens 26 - 38)	(studieren) immer mehr Jugendlichen (an der Uni)

(b) Erweiterte L2-Chunks

(1) /[Mm]einer/ . /[Mm]eining/ . /[Nn]ach/ (128)

1) Path: FalkoEssayL2WHIGv2.0 > BNG-2010-03-050 (tokens 542 - 559)

Jedoch	meiner Meinung nach	ist es noch sehr wichtig
--------	---------------------	--------------------------

2) Path: falkoEssayL2v2.4 > fu128_2006_10c_L2v2.4 (tokens 498 - 515)

	Meiner Meinung nach	sind einige Vertretern akademischer Feminismus
--	---------------------	------------------------------------------------

3) Path: falkoEssayL2v2.4 > fkb017_2008_07_L2v2.4 (tokens 419 - 436)

dass die Frauen	meiner Meinung nach	dem alten dänischen Sprich- wort folgen sollten
-----------------	---------------------	-------------------------------------------------

4) Path: falkoEssayL2v2.4 > fkb040_2008_08_L2v2.4 (tokens 20 - 37)

	Meiner Meinung nach	hängt es davon ab,
--	---------------------	--------------------

5) Path: falkoEssayL2v2.4 > hu004_2006_09_L2v2.4 (tokens 312 - 329)

, aber	meiner Meinung nach	, ist der freie Markt das einzige Manier
--------	---------------------	------------------------------------------

(2) /[li]ch/ . /[Gg]laube/ . ", " . /[Dd]ass/ (36)

1) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-156 (tokens 634 - 652)

Aber	ich glaube, dass	es möglich ist
------	------------------	----------------

2) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-097 (tokens 624 - 642)

	Ich glaube, dass	für jedes Problem es eine Lösung gibt
--	------------------	---------------------------------------

3) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-166 (tokens 572 - 590)

	Ich glaube, dass	den Feminismus riesige Vorteile zu Frauen heutzutage gegeben hat
--	------------------	------------------------------------------------------------------

4) Path: falkoEssayL2v2.4 > fkb008_2007_09_L2v2.4 (tokens 110 - 128)

	Ich glaube, dass	Feminismus zeigte
--	------------------	-------------------

5) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-115 (tokens 480 - 498)

schlechte Folgen, die	ich glaube, dass	man lieber vermeiden wollen würde
-----------------------	------------------	-----------------------------------

(3) /[Aa]ber/ . /[Ee]s/ . /[li]st/ (27)

1) Path: FalkoEssayL2WHIGv2.0 > BNG2-2011-03-209 (tokens 593 - 610)

	aber es ist	auch nötig um dauerhafte Änderung durchzuführen
--	-------------	-------------------------------------------------

2) Path: falkoEssayL2v2.4 > hu007_2006_09_L2v2.4 (tokens 426 - 443)

	aber es ist	trotzdem wichtig dafür zu kämpfen
--	-------------	-----------------------------------

3) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-107 (tokens 663 - 680)

	Aber es ist	auch wichtig zu sagen
--	-------------	-----------------------

4) Path: falkoEssayL2v2.4 > fkb001_2007_09_L2v2.4 (tokens 514 - 531)

	aber es ist	am Ende immer wertvoll
--	-------------	------------------------

5) Path: falkoEssayL2v2.4 > fkb061_2008_08_L2v2.4 (tokens 828 - 845)

	aber es ist	auch keine sofortige Lösung
--	-------------	-----------------------------

(4) /**[Ee]**s/ . /**[li]**st/ . /**[Aa]**ber/ (15)

1) Path: falkoEssayL2v2.4 >

nz002_2007_06_L2v2.4 (tokens 205 - 217)

	es ist aber	noetig.
--	-------------	---------

2) Path: falkoEssayL2v2.4 >

fkb062_2008_08_L2v2.4 (tokens 357 - 369)

	es ist aber	notwendig, dass
--	-------------	-----------------

3) Path: FalkoEssayL2WHIGv2.0 > BNG-

2010-03-031 (tokens 272 - 284)

	Es ist aber	wahr, dass
--	-------------	------------

4) Path: falkoEssayL2v2.4 >

fkb014_2008_07_L2v2.4 (tokens 216 - 228)

	Es ist aber	klar, dass
--	-------------	------------

5) Path: falkoEssayL2v2.4 >

fkb033_2008_07_L2v2.4 (tokens 389 - 401)

	Es ist aber	nicht der Fall, dass
--	-------------	----------------------

(5) /**[Ff]**ür/ . /**[Mm]**ich/ . /**[li]**st/ . /**[Ee]**s/ (3) & /**[Ff]**ür/ . /**[Mm]**ich/ . ", " . /**[li]**st/ . /**[Ee]**s/ (2)

1) Path: FalkoEssayL2WHIGv2.0 > BNG-

2010-04-084 (tokens 263 - 276)

	Für mich ist es	hier, dass wir unsere Problem finden.
--	-----------------	---------------------------------------

2) Path: FalkoEssayL2WHIGv2.0 >

BNG2-2010-11-101 (tokens 589 - 602)

	Für mich ist es	wichtig, dass Kinder
--	-----------------	----------------------

3) Path: falkoEssayL2v2.4 >

fkb034_2008_07_L2v2.4 (tokens 479 - 492)

ob das Ende feministisch klingt oder mehr anti-feministisch:	für mich ist es	aber schon feministisch.
--------------------------------------------------------------	-----------------	--------------------------

4) Path: falkoEssayL2v2.4 >

nz002_2007_05_L2v2.4 (tokens 115 - 129)

	Für mich, ist es	nicht schwierig eine Frau zu sein, weil
--	------------------	-----------------------------------------

5) Path: FalkoEssayL2WHIGv2.0 > BNG-

2010-03-058 (tokens 273 - 287)

Auf der anderen Seite,	für mich, ist es	klar, dass
------------------------	------------------	------------

(6) /**[Ee]**s/ . /**[li]**st/ . /**[Nn]**icht/ . /**[Nn]**ur/ (7)

1) Path: FalkoEssayL2WHIGv2.0 > BNG-

2010-03-021 (tokens 528 - 546)

	Es ist nicht nur	akademisch sondern auch praxisorientiert
--	------------------	------------------------------------------

2) Path: FalkoEssayL2WHIGv2.0 >

BNG2-2010-11-104 (tokens 626 - 644)

	Es ist nicht nur	die Kriminelle die leiden, aber auch die Gesellschaft
--	------------------	-------------------------------------------------------

3) Path: falkoEssayL2v2.4 >

cbs015_2007_10_L2v2.4 (tokens 338 - 356)

	Es ist nicht nur	Arbeitslosigkeit, die das Leistungsvermögen eines Menschen beeinträchtigen kann. Viele Menschen sind
--	------------------	------------------------------------------------------------------------------------------------------

4) Path: FalkoEssayL2WHIGv2.0 > BNG-

2010-03-022 (tokens 341 - 359)

	Es ist nicht nur	, dass die Universitätsabschlüsse nicht praxisorientiert sind das problematisch ist.
--	------------------	--------------------------------------------------------------------------------------

5) Path: FalkoEssayL2WHIGv2.0 >

BNG2-2011-03-188 (tokens 552 - 570)

	Es ist nicht nur	Afrika, sondern auch Asien, wo
--	------------------	--------------------------------

(7) /[[Ee]gal/ . /[[Oo]b/ . /[[Mm]an/ (6)

1) Path: FalkoEssayL2WHIGv2.0 > BNG-2010-03-040 (tokens 494 - 511)

	egal ob man	während dieser Zeit Praktikum macht oder nicht
--	-------------	------------------------------------------------

2) Path: FalkoEssayL2WHIGv2.0 > BNG-2010-03-025 (tokens 373 - 390)

	Egal ob man	es weißt oder nicht
--	-------------	---------------------

3) Path: FalkoEssayL2WHIGv2.0 > BNG-2010-03-036 (tokens 844 - 859)

	egal ob man	immer noch nur an der Universität studiert.
--	-------------	---------------------------------------------

4) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-133 (tokens 552 - 564)

	egal ob man	ein Studium oder zehn Jahre beruflicher Erfahrung für einen Job benötigt,
--	-------------	---------------------------------------------------------------------------

5) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-134 (tokens 116 - 133)

	egal ob man	Sportler oder Sänger oder Spieler ist
--	-------------	---------------------------------------

(8) /[[Ee]s/ . /[[Kk]ommt/ . /[[Dd]arauf/ . /[[Aa]n/ (4)

1) Path: FalkoEssayL2WHIGv2.0 > BNG-2010-04-080 (tokens 76 - 89)

	Es kommt darauf an	, wie das Konzept von 'Feminismus interpretier is
--	--------------------	---------------------------------------------------

2) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-129 (tokens 443 - 456)

Persönlich, es ist meine Meinung, dass	es kommt darauf an	die Gründe für die Entlohnung.
----------------------------------------	--------------------	--------------------------------

3) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-169 (tokens 715 - 728)

(...) aber die zeigen ziemlich klar, dass	es kommt darauf an	wie man diese Aussage sieht.
-------------------------------------------	--------------------	------------------------------

4) Path: falkoEssayL2v2.4 > cbs003_2007_10_L2v2.4 (tokens 719 - 732)

	Es kommt darauf an	, was man selber im Ausbildung liegt und was man später damit macht.
--	--------------------	----------------------------------------------------------------------

(c) Vergleich L1-L2

Chunk	L1		L2	
	Freq.	Freq. pro Mio.	Freq.	Freq. pro Mio.
in den letzten jahren	13	184	14	51
in den meisten fällen	10	142	7	25
ich bin der meinung	7	99	17	62
die rolle der frau	7	99	2	7
in der heutigen gesellschaft	7	99	13	47
auf der anderen seite	7	99	47	171
in der heutigen zeit	5	71	3	11
im großen und ganzen	5	71	9	33
auf der ganzen welt	5	71	1	4

Tabelle 1. Die frequentesten viergliedrigen L1-Chunks im Vergleich zu L2.

Chunk	L1		L2	
	Freq.	Freq. pro Mio.	Freq.	Freq. pro Mio.
im großen und ganzen	5	71	9	33
auf der anderen seite	7	99	47	171
ich bin der meinung	7	99	17	62
auf der ganzen welt	5	71	1	4
in der heutigen zeit	5	71	3	11
in den meisten fällen	10	142	7	25
in der heutigen gesellschaft	7	99	13	47
in den letzten jahren	13	184	14	51
die rolle der frau	7	99	2	7

Tabelle 2. Die assoziativsten viergliedrigen L1-Chunks im Vergleich zu L2.

(d) Fehlerhafte Verwendung?

(1) /[Tt]ypen/ . "von" . pos="NN" (9)

1) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-099 (tokens 88 - 100)	viele Typen von	Diebstahl
2) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-161 (tokens 40 - 52)	unterschiedliche Typen von	Kriminellen
3) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-099 (tokens 48 - 60)	andere Typen von	Kriminalität
4) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-146 (tokens 326 - 338)	verschiedene Typen von	Leute
5) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-131 (tokens 458 - 470)	Die Forderung dieser Typen von	Arbeit

(2) /[Zz]um/ . /[Ss]chluss/ . "glaube" . "ich" (5)

1) Path: FalkoEssayL2WHIGv2.0 > BNG-2010-03-026 (tokens 540 - 553)	Zum Schluss glaube ich	, dass die Universitätsabschlüsse nicht von geringem Wert sind, ob sie praxisorientiert sind oder nicht.
2) Path: falkoEssayL2v2.4 > fkb003_2008_07_L2v2.4 (tokens 578 - 591)	Zum Schluss glaube ich	, dass der Feminismus den Interessen der Frauen sehr viel nützen könne, aber es
3) Path: FalkoEssayL2WHIGv2.0 > BNG-2010-03-041 (tokens 452 - 465)	Zum Schluss glaube ich	dass praxisorientierte lernen wichtig ist
4) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-115 (tokens 448 - 461)	Zum Schluss glaube ich	, dass Kriminalität zählt sich nicht aus.
5) Path: FalkoEssayL2WHIGv2.0 > BNG-2010-03-019 (tokens 519 - 532)	Zum Schluss glaube ich	, dass es wichtig ist viele akademische und nicht akademische Erfahrungen an der Universität zu sammeln,

(3) /[Dd]ie/ . /[Mm]ehrheit/ . "von" . pos="NN" (10)

1) Path: FalkoEssayL2WHIGv2.0 > BNG2-2011-02-177 (tokens 708 - 721)	Die Mehrheit von	Kriminellen
2) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-120 (tokens 26 - 39)	die Mehrheit von	Leuten
3) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-099 (tokens 529 - 542)	die Mehrheit von	Kriminalität
4) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-099 (tokens 544 - 557)	Die Mehrheit von	Leute
5) Path: FalkoEssayL2WHIGv2.0 > BNG2-2010-11-114 (tokens 605 - 618)	Die Mehrheit von	Verbrecher

Versicherung an Eides statt

Hiermit versichere ich an Eides statt und durch meine Unterschrift, dass die vorliegende Arbeit von mir selbstständig, ohne fremde Hilfe angefertigt wurde. Inhalte und Passagen, die aus fremden Quellen stammen und direkt oder indirekt übernommen wurden, wurden als solche kenntlich gemacht. Ferner versichere ich, dass ich keine andere, außer der im Literaturverzeichnis angegebenen Literatur verwendet habe. Diese Versicherung bezieht sich sowohl auf Textinhalte als auch auf alle enthaltenen Abbildungen und Tabellen.

Die elektronische Version der Arbeit stimmt mit der gedruckten Version überein.

Leipzig, den 13. Dezember 2016

A handwritten signature in blue ink, reading 'S. Schmidt'. The signature is written in a cursive style with a large 'S' and a distinct 'Schmidt'.

Susanne Lisa Schmidt